TEACHERS' FORUM

Asset-Oriented Approaches To Learner Corpus Data

SHELLEY STAPLES

University of Arizona
E-mail: slstaples@arizona.edu

VALENTINA VINOKUROVA

Defense Language Institute
E-mail: valentina.vinokurova@dliflc.edu

ASYA GORLOVA

University of Arizona
E-mail: agorlova@arizona.edu

MARIANA CENTANIN-BERTHO

Yale University
E-mail: mariana.centaninbertho@yale.edu

BRUNA SOMMER-FARIAS

Michigan State University
E-mail: fariasbr@msu.edu

ALEKSEY NOVIKOV

Defense Language Institute
E-mail: aleksey.novikov@dliflc.edu

In this article, we discuss how learner corpus data can be used to promote asset-oriented approaches to language learning. We discuss how four tenets of asset-oriented approaches—challenges to the native speaker norm, accessibility/authenticity, advocacy, and agency—can be encouraged through using learner tasks. We introduce specific activities for Portuguese and Russian classrooms to promote this approach, which are freely available through our blog, and provide preliminary results from our teacher and student feedback on these activities.

INTRODUCTION

Learner corpora, which are collections of texts formed from learner spoken (transcribed) and written tasks, have been recognized as an important resource for language teachers (Granger, 2024). However, error analysis and correction have traditionally played a major role in learner corpus research (see, for example, an overview in Götz & Granger, 2024). Consequently, these topics have dominated learner corpus-based pedagogical approaches as an effective way to improve students' accuracy (Götz & Granger, 2024).

In this article, we describe an alternative, asset-oriented perspective on learner corpora and corpus-based pedagogy that speaks to trends in research and practice in language teaching and broader educational contexts; these involve an intentional shift toward more student-centric, culturally relevant, and critical pedagogies. Simply put, asset-oriented perspectives involve a shift from deficit-oriented approaches to learner language, focusing on what learners can do rather than focusing solely on errors or lacks (Soto, 2021). More specifically, we take up Shapiro's (2022) "both/and" approach, which acknowledges the need to critique standard language ideologies and value learner contributions while still exposing students to pathways for successful communication and language development. We see learner tasks (spoken and

written) as "allowable contributions to the genre[s they are writing or speaking in]" (Tribble, 2001) and align with Fox and Chang-Bacon's approach to error as part of learners' developmental processes rather than seeing learner language as a "problem to be fixed" (2023, p. 3). Our work is based on the Multilingual Academic Corpus of Assignments-Writing and Speech (MACAWS) corpus (Staples et al., 2019-). This corpus consists of students' written and transcribed audio/video tasks, produced as part of the curriculum for Portuguese and Russian beginner to intermediate world language classrooms.

HOW CAN LEARNER CORPORA BE USED IN LANGUAGE CLASSROOMS?

Learner corpora, like L1 corpora, can be used in the classroom via data-driven learning (DDL). The basic idea behind DDL is to give learners access to data through guided discovery tasks that promote inductive approaches to language learning, where learners discover lexico-grammatical patterns rather than learning deductive grammar rules (Johns, 1990). DDL can include both learners' direct access to corpora (unmediated DDL) or indirect access (mediated), meaning that teachers select output and prepare materials based on corpora. Mediated DDL has been the most common approach to using corpora reported in the literature, as teachers have expressed the need to mediate instruction to counterbalance technological accessibility issues and to train students to use corpus tools (Götz & Granger, 2024). Drawing from this approach (mediated DDL), the activities we present here are designed to alleviate the need for extensive corpus knowledge or training on the part of the instructor and students. However, we also provide teacher notes in many of the activities as well as basic background on the corpus and corpus-based instruction in a webinar series that teachers can access prior to using our materials.

Within the existing DDL literature on learner corpora, it is common to compare learner corpora with expert-written corpora (such as journal articles). However, the latter often contain language quite far from what is expected of learners at a particular proficiency level. Comparing learner and expert writing often leads to a discussion where learner language is not used as an example of development but as a stage to overcome. For instance, the results of such comparisons refer to learner language patterns as 'overuse' or 'underuse' of features in comparison to the expert corpora (Granger, 2015). Though the terms are meant to be descriptive and not evaluative, they still position expert language use (often also primarily produced by native speakers) as the baseline. In our view, this position may hinder a more positive application of learner tasks.

DEFINING AN ASSET-ORIENTED APPROACH TO LEARNER CORPORA

So, what is an asset-oriented approach to learner corpora? There are a number of perspectives in the literature on multilingual learners that promote asset-oriented approaches, including Culturally Responsive Pedagogy, Critical Language Awareness, Critical Multilingual Language Awareness, Linguistically Responsive Instruction, and Translanguaging Pedagogy (Burton et al., 2024; Flint & Jaggers, 2021; García & Li, 2014; Lucas & Villegas, 2013; Bucholtz et al., 2014; Paris & Alim, 2017; Shapiro, 2022; Simonsen, 2019). These pedagogies share several tenets that can be used to outline the overarching theme of asset-based approaches, including asset-based approaches to learner corpora. In addition, we want to clarify that engaging with

learner corpora in the ways we outline here is only one aspect of shifting pedagogy to an asset-oriented approach. Others have written more extensively on how such approaches can be used to transform curricula (see, e.g., Bucholtz et al., 2014). We hope that our suggestions add to the range of activities proposed by scholars and practitioners to bolster such an approach. We want to make clear, however, that simply using learner corpora does not equate with an asset-oriented approach but rather requires teachers to engage with asset-oriented principles, such as the ones described below.

Simply put, asset-oriented perspectives involve a shift from deficit-oriented approaches to learner language (Soto, 2021). While the pedagogical approaches listed above differ in their engagement with students' cultural, social and linguistic backgrounds and social justice goals, they all share a focus on highlighting learners' strengths, which we see as integral to our approach. While we do not suggest that learner errors never be addressed in the classroom, we argue that there are many positive examples to be drawn from the language learning process, and that we can use these to bolster language learning in multilingual environments. In particular, we see four central tenets of asset-oriented pedagogy as relevant to the use of learner corpus tasks: challenges to the native speaker norm, accessibility/authenticity, advocacy, and agency.

One major principle of asset-oriented approaches involves a challenge to the native speaker norm, or "standardized" language use (Fox & Chang-Bacon, 2023; Shapiro, 2022; Soto, 2021). In our use of learner corpora as models for other learners, asset-oriented approaches challenge the native speaker model in two key ways. First, we value the developmental stages of the learners in their own right, rather than requiring a comparison with the native speaker norm to make their language meaningful for the classroom. Students can see themselves represented and legitimized more readily through the language used by other learners, and thus understand that they are on a developmental path rather than simply not meeting the standards of a native speaker norm. In addition, we see this approach as aligning with translingual pedagogies that support learners' use of multilingual repertoires to achieve communicative goals (García & Li, 2014; Gilquin, 2022).

By focusing on the latter and by including authentic corpus samples that are contextualized by purpose (we identified the metagenre and topic for all tasks in the corpus), students can see how other learners have successfully achieved those goals, and/or evaluate their success based on factors other than accuracy. Students can be encouraged to directly discuss the use of language forms within learner tasks as potentially countering a standardized norm while working towards task completion for a specific proficiency target.

Somewhat related to this factor, we see learner tasks as providing more accessible forms of language and more appropriate contexts of language use for learners, particularly those at lower levels of proficiency. These tasks are also uniquely positioned to bring in language and culture that is relevant, and thus more "authentic" for students. As Seidlhofer (2002) indicated, learner tasks contain language that is familiar and nonthreatening. Many existing L1 corpora contain language that students may struggle with or not relate to, to the point that they are inaccessible. In addition, learner tasks are from contexts familiar to students (e.g., the types of assignments produced by learners are typical of communicative situations found in language learning classrooms, unlike the communicative situations found in expert corpora). In this way, using learner corpora as both language and content for the language learning classroom responds to Soto's (2021) call for content that supports students based on their cultural and linguistic backgrounds.

Another aspect of asset-oriented pedagogy is a focus on student advocacy. Bucholtz et al. (2014) outline several goals as part of their efforts at sociolinguistic justice, one of which

is "To promote the validity of one's own and others' full linguistic repertoires for symbolic and/or communicative use in a wide range of social spheres, including not only the intimate and informal settings of home and community but also formal, public, and institutional settings." (p. 146). We see our use of learner language as acknowledging students' right to use their own language (c.f. SRTOL from composition studies; CCCC, 1974). This advocacy involves "locating students" (King & Swartz, 2016, cited in Flint & Jaggers, 2021), in other words, centering students' lives, languages, and experiences in the classroom (Flint & Jaggers, 2021, p. 259). In this way, uses of learner corpora that reflect students' language choices can be seen as an aspect of student-centered, asset-based learning.

Finally, and related to advocacy, student agency posits students as possessing expertise in articulating their own experiences and ways of understanding the world, determining what language is appropriate for a given context, and using language of their choice. Bucholtz et al. (2014) articulate this goal as promoting students' ability to approach their use of language as linguistic experts and critical thinkers. This perspective clearly runs counter to a deficit approach to learner language. Related to this concept is the notion of rhetorical agency (Shapiro, 2022). This approach advocates for students making informed choices about their language use, based on their understanding of the audience, purpose, and other factors that influence what is known as the "rhetorical situation" (or a communicative situation of language use). The concept of rhetorical agency also allows teachers to bring standard language use into discussion with student tasks, providing a space for dialogue as students make decisions about whether or not to align with standardized patterns in purposeful ways. Importantly, learner language corpora, like most language corpora, contain examples of both standard and nonstandard use that provide opportunities for learners to gain language awareness and rhetorical agency through class discussions.

OUR EFFORTS AT MORE ASSET-ORIENTED APPROACHES

Our efforts towards more asset-oriented approaches use the MACAWS corpus (Staples et al., 2019). The assignments used to build this corpus (from Russian and Portuguese, to be expanded to Spanish and Chinese in the future) represent authentic language from class assignments at beginning and intermediate levels of instruction. The corpus contains (transcribed) spoken and written tasks on topics such as family, daily routine, and future plans, and currently contains over 500,000 words of Portuguese and around 240,000 words of Russian (see Sommer-Farias et al., 2022 for a description of the corpus as well as our transcription and classification methods). Positioning these learner classroom tasks as authentic supports a view of the classroom as a "legitimate sociocultural context in which meaningful communication takes place" (Simonsen, 2019, p. 248). The materials we create from the corpus have a similar context in mind: beginning to intermediate learners in college-level world language classrooms.

We designed our activities using principles of corpus-based instruction that are usually reserved for L1 speaker corpora (see, e.g., Reppen, 2010; Friginal, 2018), but that can be applied to L2 corpora to encourage asset-based approaches. These principles, coupled with learner data from our corpus, support our asset-oriented approach. First, the inductive approach of data-driven learning—in which students are engaged as "language detectives"—is student-centered and supports agency in the learning process (Johns, 1997). By using learner language in this approach, it is legitimized as an object to study and learn from—importantly, not as a negative example (as in error analysis), but as offering models. Second, in corpus-

based instruction, students can be introduced to concepts that challenge the idea of a "standard" or "neutral" language that is invariable. By including learner data in instruction, students are exposed to developmentally appropriate language, variations across assignments that are relevant to their learning process, and trained to register differences (e.g., between speech and writing).

Teachers may be concerned about introducing errors into their classrooms through learner tasks. However, materials created from the corpus show that students, even at beginning levels, mostly provide positive examples of linguistic phenomena, with frequencies of occurrence guiding the language included in material design. Since the tasks are at developmentally appropriate levels, students engaging with the corpus-based materials can see how other students performed similar tasks successfully, as well as how they use developmentally appropriate language in their assignments. In this way, we validate the students' stage of their developmental journey, which doesn't negate the fact that they are still learning. We are also able to introduce more advanced language choices, for example by highlighting students' use of more varied vocabulary, or by sharing tasks from a slightly higher developmental level.

Examples from Russian

A few examples of activities we have developed for learners of Russian help demonstrate our asset-based approach in practice. Our first example is a lesson that introduces students to the features of spoken language, particularly spontaneous or improvised speech.[1] On the one hand, this lesson prepares our students to work with speech samples in future lessons. On the other hand, it advocates for students' own language and challenges the native speaker ideal. Specifically, the very first activity in the lesson invites students to examine nine short samples of native (NS) and non-native transcribed speech (NNS) and to notice the occurrence of hesitation markers in all samples, regardless of who produced them.[2] While NS-NNS is not a distinction that we are interested in promoting, we work with concepts that are at the forefront of dominant language ideologies in order to challenge and eventually deconstruct them. As such, through a series of activities in this lesson, we guide students through identifying hesitations, false starts, repetitions, and self-repair in speech samples and help them notice that they are typical of spontaneous speech produced by both native and non-native speakers. Even though the usage pattern of those features might be different in NS and NNS speech, the focus is on the function of such features, e.g., repairing a linguistic mistake, emphasizing a linguistic repair, or giving yourself time to think of a more accurate word. By analyzing the function of spontaneous speech features, students are less likely to feel anxious about the presence of those features in their speech, and can potentially use them more strategically.

Validation of language produced by students is a common thread across our lessons, as all of them utilize tasks produced by other learners as models. In addition, many of our lessons focus on exploring topics, genres, and types of assignments that are particularly relevant to our students and their academic context. For example, as part of the module on food in our university's Russian language program, students are often asked to create recipes and record videos of following them. To assist learners in this task, we have designed a lesson in which they can explore the genre of video recipes by referring to a sample recipe produced

[1] We were inspired by the Corpus for Schools project from Lancaster University for this activity.
[2] The native speaker in this corpus sample was the teacher. The sample was drawn from an exam, which is why there are native and non-native speech examples in this sample.

by a student in the corpus. By using a sample produced by a learner, we aim to not only validate learner language but also present students with an accessible authentic task produced in a relevant context. In a series of activities within this lesson, students answer comprehension questions, examine the structure of a video recipe, notice and learn more about the language typical of the genre, and evaluate the overall effectiveness of the sample recipe.

Examples from Portuguese

The activities created using L2 Portuguese learner corpus data also position learners as legitimate speakers and center their lives, languages, and experiences to introduce and contextualize language use in situations that are familiar to them. For example, in a beginner-level activity created to teach several uses of the verb "ficar" ("to get"), examples from the corpus describe situations that are relatable to students, thus making the different functions of the verb easier to process. The activity is inductive, as it asks students to notice patterns associated with the verb "ficar". The verb co-occurs with words that are accessible and familiar at their proficiency level, such as "fico feliz" ("I get happy") and "fiquei no hotel" ("I stayed at the hotel"). These examples occur within topics pertinent to their daily lives and discussed in the curriculum, making it easier to inductively learn and discuss the pattern with classmates describing their own experiences within similar realms (travel, home, or feelings).

Responses From Instructors And Students

Alongside corpus building, our team has offered professional development workshops for teachers in which we show teachers how to use our corpus, ready-made materials we developed based on our data, and talk about how corpus pedagogies can make language learning more learner-centered and asset-based (see our blog for recorded workshops). Through these workshops and from post-implementation surveys, we have received positive feedback supporting the benefits that an asset-based approach to learner corpora can bring to both students and instructors.

Instructors have reported that our activities prompt students to look at both the language they are learning and how they are using it in a new light, leading to them acquiring new skills. Commenting on the activity described above introducing the features of spoken language, an instructor shared, "Students were reassured that disfluencies happen in everyone's speech, and our goal is not to get rid of them but embrace them and be comfortable with them." This demonstrates that instructors perceive the importance of validating students' language use and resisting standard language ideologies that permeate language learning contexts. Additionally, instructors express that engaging with spoken corpus activities provides students with important cultural experiences of engaging with the variety of ways the language can be used by their peers. Both students and instructors point out that tasks are accessible, relevant to their learning contexts, and connected to the real world and students' lived experiences. In particular, students reported that they "like getting to identify the patterns like a puzzle" and "like seeing examples from other students."

Instructors also note that learner corpus activities contribute to students' development of language awareness. By seeing the linguistic and discourse patterns and differences in corpus examples, students not only learn how to create connections between meaning, form, and use, but also gain awareness of language use across contexts. This can contribute to their engagement with other linguistic varieties and demonstrate the legitimacy of learner language

and use. Instructors have also indicated that students are able to work on corpus activities independently and discover linguistic knowledge for themselves, promoting their autonomy and agency in the classroom and beyond. These comments illustrate that students get similar outcomes from a learner corpus as they may be able to get from an L1 corpus.

However, both teacher and student survey responses indicate that L2 corpora are still sometimes viewed in the classroom through the lens of deficit perspectives and standard language ideologies. Importantly, only one student out of 134 surveyed found it problematic to use student tasks in their instructional context (Portuguese or Russian). It should be noted, nonetheless, that a few students still seemed to focus to some extent on grammatically correct and native-like language use. For example, one student indicated, "I liked that it was a spontaneous activity and challenged me to use unprepared speech! I feel that with more practice in this specific activity, my unprepared speech will come more natural and grammatically correct." While this is not necessarily a problematic goal, it does seem to reflect underlying ideologies about the purpose of language learning and the views of language that permeate language classrooms. Combatting language ideologies in the classroom can be seen as one of the goals of asset-based pedagogies: as the first step, it might be helpful for teachers to directly address students' concerns about "errors" as they are working with learner corpora.

IMPLICATIONS AND CONCLUSIONS

In this short report we have attempted to argue for an asset-oriented approach to the use of learner corpora in language classrooms, providing concrete examples of how learner corpora can be used for this purpose. We also provided support for this approach from the feedback we have received from teachers and students who have engaged with our asset-oriented materials developed from the MACAWS corpus.

We do wish to acknowledge the challenges teachers may face in designing and implementing asset-oriented materials from learner corpora. First, teachers may not be in a language program with an administrator who sees the value of these approaches, and it may be difficult to deviate from a preexisting curriculum to include a focus on these types of activities. In the creation of our materials, we did look closely at textbooks and assignments used in our own teaching context to find places where the learner corpus data could complement language found in the existing curriculum. We hope that our examples of activities show how these can be integrated into classrooms without a complete curricular overhaul, although of course more extensive revisions with asset-oriented approaches in mind (e.g., to rubrics, learning outcomes, etc.) would create more extensive opportunities to further this approach (see e.g., Shapiro, 2022; Beaudrie et al., 2021 for examples of revamping curricula to align with asset-oriented principles).

Another challenge teachers may face is getting access to learner corpora. While there are a number of resources in English, there are fewer for less commonly taught languages, which was one of our purposes in building the MACAWS corpus. In the future, our corpus will include tasks from Spanish and Chinese learners, using the same model as that used for the Portuguese and Russian corpora. Readers are also referred to the extensive list of learner corpora that can be found on our blog. We also want to point out that use of learner texts does not have to mean use of a large-scale corpus. Teachers who have access to student samples can consider using those samples in similar ways to what we describe here.

Creating these types of materials can be a heavy load for teachers, even those with a great deal of teaching experience. Teachers will need to utilize their pedagogical knowledge of

their teaching context but also be comfortable with the content knowledge needed to determine relevant language to draw from the corpus, the technological knowledge needed to navigate it, and the knowledge needed to combine all of this into an effective lesson (Technological Pedagogical Content Knowledge or TPACK; Koehler & Mishra, 2005). We hope that our materials, found at our blog, can be helpful both for providing activities that can be directly applicable to Portuguese and Russian language classrooms and also to serve as examples of what such materials can look like outside of these contexts. We have been pleased to see that instructors outside of our institution have been able to use the corpus-based materials provided on our blog with little to no corpus training. We also provide a webinar series introducing the corpus and the principles behind our materials development, as well as an example of how these materials can be implemented in an actual language learning classroom.

Finally, we wish to acknowledge that it will be important to measure the impact of these materials through more than teacher and student perceptions about their learning. In the future, we do plan to conduct an analysis of tasks produced by students with and without these corpus-based materials.

REFERENCES

Beaudrie, S., Amezcua, A., & Loza, S. (2021). Critical language awareness in the heritage language classroom: Design, implementation, and evaluation of a curricular intervention. *International Multilingual Research Journal, 15*(1): 61-81.

Bucholtz, M., Lopez, A., Mojarro, A., Skapoulli, E., VanderStouwe, C., & Warner-Garcia, S. (2014). Sociolinguistic justice in the schools: Student researchers as linguistic experts. *Language and Linguistics Compass, 8*(4), 144–157.

Burton, J., Wong, W., & Rajendram, S. (2024). Translanguaging for critical multilingual language awareness: preparing teacher candidates to support multilingual learners in classrooms. *International Multilingual Research Journal*, 1-24.

Centre for English Corpus Linguistics (accessed January 14, 2025): Learner Corpora around the World. *Louvain-la-Neuve: Université catholique de Louvain*. https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html

Conference on College Composition and Communication. (1974). Students' right to their own language. *College Composition and Communication, 25*(3): 1-32.

Flint, A.S. & Jaggers, W. (2021). You matter here: The impact of asset-based pedagogies on learning. *Theory Into Practice, 60*(3), 254-264.

Fox, W. & Chang-Bacon, C. (2023). Critical language awareness through advocacy: Disrupting deficit discourse through simulation in L2 writing teacher education. *Journal of Second Language Writing, 60*.

Friginal, E. (2018). *Corpus linguistics for English teachers: Tools, online resources, and classroom activities*. New York: Routledge.

García O. & Li W. (2014). *Translanguaging: Language, Bilingualism and Education*. London: Palgrave Macmillan Pivot.

Gilquin, G. (2022). Translanguaging and data-driven learning: How corpora can help leverage learners' multilingual repertoires. TESL-EJ, 26(3), available at: https://doi.org/10.55593/ej.26103a22

Götz, S., & Granger, S. (2024). Learner corpus research for pedagogical purposes: An overview and some research perspectives. *International Journal of Learner Corpus Research, 10*(1), 1-38.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7-24.

Granger, S. (2024). From early to future learner corpus research. *International Journal of Learner Corpus Research, 10*(3): 247 - 279.

Johns, T. (1998). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Routledge.

King, J. E., & Swartz, E. (2016). *The Afrocentric praxis of teaching for freedom*. Routledge.
Koehler M.J. & Mishra P. (2005) What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research, 32*, 131–152.
Lucas, T., & Villegas, A.M. (2013). Preparing linguistically responsive teachers: Laying the foundation in preservice teacher education.*Theory Into Practice, 52*, 98–109.
Paris, H.S. & Alim, D. (2017). What Is culturally sustaining pedagogy and why does It matter? In H.S. Paris & D. Alim (Eds.), *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world* (pp. 1-24). Columbia Teachers College Press.
Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
Shapiro, S. (2022). *Cultivating critical language awareness in the writing classroom*. Routledge.
Simonsen, R. (2019). An analysis of the problematic discourse surrounding "authentic texts". *Hispania, 102*(2), 245-258.
Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning drivendata. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213–234). Philadelphia, PA: John Benjamins.
Sommer-Farias, B., Novikov, A., Picoral, A., Bertho, M., & Staples, S. (2022). A multilingual learner corpus for less commonly taught languages. *International Journal of Learner Corpus Research, 8*(2): 261-282.
Soto, C. S. (2021). Asset-based pedagogies for English learner curriculum. *Imagine: A Promise Scholars and McNair Scholars Journal.* Available at: https://imagine.sa.ucsb.edu/issue/25/2023/asset-based-pedagogies-english-learner-curriculum
Staples, S., Novikov, A., Picoral, A., Sommer-Farias, B. (2019-). *Multilingual Academic Corpus of Assignments - Writing and Speech*. Available at https://macaws.corporaproject.org/