# Collaborative Commentary for

# Understanding Communication Disorders

Brian MacWhinney

Davida Fromm

John Kowalski

Department of Psychology, Carnegie Mellon University, Pittsburgh PA

Correspondence to Brian MacWhinney: macw@cmu.edu

# Abstract

**Purpose**: The goal of the Collaborative Commentary (CC) system is to make the TalkBank adult clinical databases – including AphasiaBank, DementiaBank, RHDBank, and TBIBank -– open to commentary and analysis from the full community of researchers, instructors, students, and clinicians.

**Method**: CC allows a group leader to establish a commentary group and invite colleagues or students to join as members of the group. Members can then browse through the transcript database using the TalkBankBrowser (TBB). When they wish to insert a comment, they click on the utterance line number or drag the cursor across a range of utterances and a window opens to receive the comment. The comment can include open text along with codes selected from a predefined set of codes created by that commentary group.

**Results**: CC was released for public use in August 2022. It is being used currently in five research projects and eight classes. An important feature of CC is its ability to evaluate the reliability of coding systems and to sharpen analytic categories. By familiarizing instructors and researchers with the capabilities of CC, we expect to see an increasing usage of CC for a variety of clinical and research applications.

**Conclusions**: CC can contribute to a better understanding of connected speech features in aphasia, dementia, RHD, and TBI. CC represents an extreme innovation not only for the study of adult neurogenic communication disorders, but for the study of spoken language generally.

# Introduction

Collaborative Commentary (CC) is a tool that allows groups to collaboratively code and comment on transcripts in the TalkBank databases. TalkBank is the largest publicly available source of data on spoken language interactions. All the data in TalkBank are coded in a consistent format and the transcripts and media can be played back directly through the TalkBank Browser. This open accessibility makes TalkBank data an excellent target for collaborative commentary and analysis.

Using CC, researchers, clinicians, and students can access video recordings of spoken language interactions in the browser, watch the video, follow along with the linked transcript, and enter codes or comments directly attached to utterances in the transcript. These codes or comments are then visible to everyone in the commentary group – which may be the class, the research group, or clinical trainees. The rapid development and ubiquitous presence of broadband web connections offers enormous opportunities for collaborative learning and theory testing. No longer limited to working on isolated computer files, groups of people can now focus their mutual attention on a common set of interactions in a shared database to promote deeper and wider understandings, test theories, and elaborate analysis systems.

# History

Nearly two decades ago, MacWhinney and colleagues (2004) proposed the design of a CC system for spoken language data. However, the web tools available for spoken language at that time made development difficult. During the same period, the W3C Annotea project at [www.w3.org/2001/Annotea](http://www.w3.org/2001/Annotea) developed tools for web-based commentary on written language materials. The development of commentary tools for written materials has continued with systems such as AlvisAE (Papazian et al., 2012), Cromer (Girardi et al., 2014), INCEpTION (Boullosa et al., 2018), Mimir (Tablan et al., 2015), MTAS (Brouwer et al., 2017), NeuroCurator (O'Reilly et al., 2017), and WebAnno (De Castilho et al., 2016). However, these tools only work on written documents, and most do not provide methods for group-based commentary. The Hypothesis project at <https://web.hypothes.is> provides commercial tools that can be adopted in classrooms. However, Hypothesis tools are still limited to commentary on written documents on the web.

The tradition of Qualitative Data Analysis (QDA) has also led to the development of widely used commercial systems, such as ATLAS.ti, TransAna, Lexalytics, MAXQDA, Dedoose, Quirkos, and NVivo, for commenting on files on individual computers. However, none of these systems are linked to openly shared data and most work mainly on individual files on the desktop. A review of these systems can be found at <https://www.predictiveanalyticstoday.com/top-qualitative-data-analysis-software/>

To permit web-based commentary for shared data on spoken language interactions, a fully functional system must provide these features:

1. There must be open web access to transcripts of spoken language interactions linked to media. The TalkBank CC system provides this by relying on the specially designed TalkBank Browser which follows HTML5 standards for web playback.
2. For consistent analysis, the transcripts must be in a common format that identifies the speaker and breaks up each turn into its component utterances or T-units (Foster et al., 2000). Within utterances, there must be consistent methods for transcribing features such as pauses, filled pauses, retracing, errors, and dialect forms. TalkBank CC achieves this by relying on the fact that all data in TalkBank are in the CHAT format which provides all these features, as well as methods for Conversation Analysis (CA) notation and morphosyntactic analysis.
3. The transcripts being analyzed must be linked on at least the utterance level to either audio or video media. This is important for understanding conversational, proxemic, and gestural features of the interactions. TalkBank CC provides this facility through time links from the transcripts to the media. Broadband connections allow for smooth and direct playback from a highlighted utterance in the transcript to the corresponding segment of the media. Prior to 2022, these links were created by the transcriber. Our new Batchalign system available at <https://github.com/talkbank> can insert these links automatically.
4. Data must be protected in accord with IRB requirements, GDPR regulations, and informed consent agreements. For clinical samples, this typically means that data must be password protected. The required levels of password protection will vary from corpus to corpus. TalkBank implements these protections through a system of user-based passwords and access permissions.
5. To provide flexibility, comments must be stored in a database separate from the main transcript database. TalkBank CC implements this by storing comments in a separate PostgreSQL database organized by commentary groups. For each comment, the following features are stored: file PID (permanent file ID), begin and end time of the media segment, time of comment creation, text of the comment, ID of the user creating the comment, group membership of the user, group status of the user, identity of the coding system and the codes in that system, and links to other comments.
6. The system must support the control of commentary group structure. Commentary group owners can provide members with privileges to either read and write or just read. Additionally, group owners can choose to hide or share comments and codes from individual group members to allow for independent judgments by each group member and later collective review.
7. It must be possible to search through a group’s comments on the basis of specific keywords, codes, transcripts, or group members. TalkBank CC provides this through searches controlled from the PostgreSQL database.
8. It must be possible to output the text of the transcript along with the newly attached comments to other programs for further analysis. These further analyses could involve checks for reliability between coders or statistical analysis of distributions of codes across transcripts. TalkBank DB facilitates this by allowing the user to save commented transcripts to the desktop for further analysis in R (<https://www.r-project.org>) or CLAN (<https://dali.talkbank.org/clan>).
9. To provide for replicability (Munafò et al., 2017), CC must store the state of the CHAT transcript database and the commentary PostgreSQL database for a given analysis at a given time point. TalkBank CC achieves this through version history in the TalkBank GitLab repository.
10. To provide sustainability, the system must be built on widely used, open-source tools. TalkBank CC achieves this by relying on NodeJS, JavaScript, and PostgreSQL, along with current web standards.

# Learning CC

To learn CC, it is good to begin with the screencasts at <https://talkbank.org/screencasts>. That page gives links to these 8 short screencasts:

1. CC-overview. This screencast explains how to find CC on the web, where to find the manual, how to start CC within the Browsable Database, how to add a comment to an utterance, and how to search for and create tags.
2. CC-new\_user. This shows how to register as a new user of CC.
3. CC-join\_group. This shows how to join a CC group or participate in a group of which you are already a member or owner.
4. CC-comments. This shows how to enter comments and codes into transcripts.
5. CC-search. This shows how to search for comments with specific tags, words, creators, transcripts, or groups. Searches will return matches in a window and then you click on each match to go directly to the locus of the comment in the transcript where you can replay the segment.
6. CC-contact\_user. This shows how to send email to the creator of a comment. The email includes the exact location and text of the comment. This is particularly useful for an instructor who wishes to provide feedback to a student.
7. CC-owner. This video explains various functions controlled by the group owner, such as adding and removing members or setting specific types of data access.
8. CC-manage. This shows how the group owner can manage read and write permissions for users.

Next, you can scan the illustrated, step-by-step manual at <https://sla.talkbank.org/CCmanual/> and follow through the steps of creating an account, choosing a password, and responding to an email that verifies your account. For practice, you may want to create your own new code along with tags for codes that interest you. The manual explains the processes of tag set creation, comment insertion, and comment searching.

# Interface Illustrations

This section provides illustrations of a few of the screens used by CC. Figure 1 shows the top-level CC page -- <https://talkbank.org/CC/> -- with links to the TalkBank Browser, the manual, screencasts, and applications of CC. Figure 2 displays the page at <https://sla.talkbank.org/TBB> which you can use to start the TalkBank Browser and the CC process. The two screenshots in Figure 3 show the dialogs that help you join a group.

# Applications

CC can be used in a wide variety of ways. This section describes some of these uses.

## Teaching

First, we illustrate how CC can be used in classroom teaching. To begin, the instructor creates a new CC group for the class, along with a set of codes or tags that students will use. As group owner, the instructor can then invite the students to be members by entering their email addresses. Here are four examples that illustrate ways to use CC for teaching.

Students learning about aphasia can be asked to identify and code behaviors such as paraphasia (semantic, phonemic, mixed), circumlocution, agrammatism, paragrammatism, jargon, anomia, empty speech, conduit d’approche, stereotypy, and perseveration. These features are coded in a tag set created by the instructor which students can use to mark each case consistently. Along with the tags, they can add comments about utterances.

Students in classes on acquired adult communication disorders can be asked to look for and comment on a participant’s ability to comprehend, self-monitor, and self-correct – all important considerations for treatment planning. They can identify which cases may demonstrate coexisting apraxia of speech and code the relevant behaviors that led them to that diagnosis. Students can practice coding correct information units (CIUs), a frequently used outcome measure in aphasia assessment (Bryant et al., 2017). At the macrostructural level, students can code features of discourse such as global coherence, cohesion, and story grammar.

Students can be asked to examine Case #2 in the RHDBank Grand Rounds tutorial at https://rhd.talkbank.org/education/class-rhd and respond to these questions: *How would you judge Phil’s stroke story? Did it embody some of the characteristics of right hemisphere discourse? If so, which ones – information content, organization, coherence, prosody, etc.? Would you agree that Phil’s Cinderella story could be described with these terms – verbose, fluent, intelligible, confabulatory, tangential? What other terms might you use instead of or in addition to those terms?*

Elise Elbourn created a fourth example assignment in conjunction with the TBIBank Grand Rounds at <https://tbi.talkbank.org/education/class-tbi>. This assignment uses Video 6a from a 56-year-old woman who presents with aphasia in addition to a cognitive communication disorder resulting from a motor vehicle accident. Elbourn told her students to read the Grand Rounds material for Video 6a, which gives a brief description of the case and poses this question: *What features of her spoken discourse are more consistent with aphasia vs cognitive-communication disorder?* Figure 4 shows the comments and codes Elise entered in the CC Video 6a transcript in piloting this exercise. With tags she had created, she identified instances of anomia and semantic paraphasias, providing evidence consistent with aphasia. She then used comments to mark instances where the speaker demonstrated difficulty understanding the main idea of the story and made an irrelevant comment, providing evidence of a cognitive communication disorder.

## Learning Clinical Practices

CC can be used for learning clinical practices. Group members can observe the proper administration and scoring of formal and informal tests. For example, AphasiaBank corpora include recorded and transcribed material from the short form of the Boston Naming Test (Kaplan et al., 2001), the Verb Naming Test from the Northwestern Assessment of Verbs and Sentences (Cho-Reyes & Thompson, 2012), the Quick Aphasia Battery (Wilson et al., 2018), the Famous People Protocol (Holland et al., 2014), and the picture description from the spontaneous speech segment of the Western Aphasia Battery-Revised (Kertesz, 2007).

To learn clinical interaction skills, students can observe a skilled clinician and then use tags to identify specific behaviors, such as comm-conv (conversational comment), comm-impr (comment on improvement), cues (provides cues), hum (uses humor), interp (interprets, restates), ques-c (closed-ended question), ques-o (open-ended question), reinf (reinforce), sugg (suggest), supp (support, sympathize), and time (slow pace). Figure 4 presents a sample of the use of these tags.

## Research Projects

Research projects that involve coding or scoring behaviors can use CC in several ways. CC allows research groups to work together asynchronously on shared materials. For example, CC makes it possible for several research assistants to be engaged in parallel at the work of coding CC transcripts. People can work at home or in the office on their own schedule and the results will be stored in a single consistent form for later analysis.

Coding systems can be developed for gesture, fluency, agrammatism, paragrammatism, main concepts, global coherence, conversational sequencing, and more. The research team can compute the interrater reliability of code use to determine if the coding system needs further development.

If coders did not establish adequate reliability on the first pass, the system can be refined by modifying the training or the codes themselves. Additional trials can be used to establish coding reliability scores right within the CC tool itself.

A research group in Australia is currently using a set of macrostructure codes to evaluate discourse on a specific task in a subgroup of the TBIBank Togher protocol database (Zhang, personal communication).

Group members may come from different perspectives that yield contrasting and conflicting interpretations of conversational and linguistic features. These differences can be most clearly studied by placing their comments and analyses in parallel on specific utterances.

* Some research projects require that each utterance be assigned a tag from a pre-specified tag set for things like speech acts, main concepts, and rhetorical features.. The CLAN program made this type of exhaustive coding possible through use of a system called Coder’s Editor. We plan to replicate this same functionality in future versions of CC.

## Automatic Tagging

The next three applications of CC rely on ongoing programming developments that are not yet complete. We discuss these here because we expect to have these functionalities available soon based on current funding.

TalkBank transcripts can be elaborated with codes entered from automatic tagging systems. One example of this is CLAN's MOR system for automatic tagging of morphological and syntactic dependency structure. Because morphosyntactic codes are so important for computing clinical profile measures such as IPSyn (Scarborough, 1990), C-NNLA (Thompson et al., 1995), DSS (Lee, 1966), C-QPA (Rochon et al., 2000), EVAL (Forbes et al., 2012), and FluCalc (Bernstein-Ratner & MacWhinney, 2018), the codes produced by MOR are integrated directly into the main transcript database on the %mor and %gra lines. Another type of automatic tags that are integrated into the database are the word-level and utterance-level time codes created by use of the Batchalign program.

There are also special purpose, automatically generated codes that should be kept separate from the main database. For example, coding for emotion terms using the LWIC system (Pennebaker et al., 2001) or coding for conversational moves using Rhetorical Structure Theory (RST) (Mann & Thompson, 1992) can be stored in segments of the CC database. Similarly, the utterance-level tags for DementiaBank transcripts created by participants in the INTERSPEECH ADReSS challenges for 2020 and 2021 (Haider et al., 2019; Luz et al., 2021) can be stored in the CC database.

A related method for adding codes to TalkBank files involves reformatting codes contributed by users that were not entered directly into the CC database.For example, we are now in the process of incorporating codes for conversational moves in the DementiaBank database contributed by Shahla Farzana (Farzana & Parde, 2022; Farzana et al., 2020) .

## Coding Untranscribed Media

Currently, data in the clinical banks have all been fully transcribed. Typically, these involve recordings of sessions lasting less than one hour. However, researchers are also interested in working with recordings taken across much longer timescales. For example, the audio recordings in HomeBank at <https://homebank.talkbank.org> each last for a full 16-hour day. Audio and video recordings of this type will become more common in the future. To use CC with this type of daylong media, a CLAN command can create a CHAT file composed of regular time segments for display in CC on the web. Researchers can then listen through the media and enter codes for specific segments without having to create a full transcript. In this way, CC provides functionality much like that found in systems such as The Observer XT at https://noldus.com.

## Citizen Science

The Zooniverse system at https://www.zooniverse.org has organized nearly 500 projects in which citizens use web interfaces to register observational, experimental, and analytic data. Using CC, we can open up deidentified data for crowd-sourced transcription, commentary, and analysis. Citizen input through CC could be particularly useful for untranscribed daylong recordings of the type currently found in HomeBank, but which could also be gathered for clinical participants. If citizen volunteers could go through segments of these materials to identify speakers, note distracting noises, spot code-switching, or point out periods of silence, their annotations would provide important training data for further automatic analysis of daylong transcripts. Apart from Zooniverse, Citizen Science applications can also be routed through systems such as AMT (Amazon Mechanical Turk), Qualtrics, Prolific, Pushkin, or Cloudresearch for participant recruitment, gamification, and payment.

## ASR and Online Transcription

The major barrier limiting more widespread use of language sample analysis (LSA) is the time it takes to transcribe interactions (Overton & Wren, 2014). Advances in speech recognition and web technology now make it possible to lower this barrier. Using our Batchalign system (<https://github.com/talkbank>), a clinician or researcher can send a 30-minute audio or video recording to the web for automatic speech recognition (ASR) processing and forced alignment and receive back a fairly accurate transcription within a few minutes. We are working to also make it possible to automatically contribute the result to the TalkBank database. Once in the database, the researcher and assistants will be able to play through the transcript on the web to add comments and correct residual errors.

# Conclusion

CC can contribute to a better understanding of connected speech features in aphasia, dementia, RHD, and TBI. CC represents an extreme innovation not only for the study of adult neurogenic communication disorders, but for the study of spoken language generally. Codes can be tracked across interactions and participants. The reliability of coding systems can be evaluated. Students can learn how to analyze clinical types and interactions. CC allows researchers to sharpen their coding and interpretation of language behaviors and the challenges people face during conversational interaction. The interpretation of the scope and causes of these difficulties can then inform assessment, classification, treatment., and treatment evaluation. Most importantly, CC makes both the raw data and the results of our comments and analyses fully public and replicable.

# Data Availability Statement

To obtain the username and password for access to CC and the adult clinical databases (AphasiaBank, DementiaBank, RHDBank, and TBIBank), researchers and licensed clinicians should first read and accept the TalkBank Ground Rules at <https://talkbank.org/share> and then send an email request for membership to [macw@cmu.edu](mailto:macw@cmu.edu) indicating their contact information, affiliation, and which banks interest them. Membership is usually granted quite quickly. Graduate and undergraduate students should obtain access by asking their faculty advisors to join as members.

# Ethics Statement

This study did not involve contact with human participants. The data in TalkBank have all received IRB review and approval at the contributing institutions.

# Author Contributions

BM designed the overall shape of the system. JK created CC, the TalkBank Browser, and TalkBankDB. DF created the screencasts for CC and is promoting use of the system. BM and DF contributed to the writing of this report.

# Conflict of Interest Declaration

The authors declare that they have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

# Funding

This work was supported by NSF HNDS-I grant 53286.1.1122907 for the development of TalkBank Collaborative Annotation.

# Acknowledgments

We wish to acknowledge the contributions of the many colleagues who have contributed their recordings and transcripts to the clinical components of TalkBank, including AphasiaBank, DementiaBank, and TBIBank. Without their collegial involvement in data-sharing, this type of open analysis of adult language disorders would not be possible.

**Figure 1:**

*Starting Page for Information about CC --* [*https://talkbank.org/CC/*](https://talkbank.org/CC/)

Text, timeline

Description automatically generated with medium confidence

**Figure 2**

*Starting page for the TalkBank Browser with the Collab Button*

Graphical user interface, text, application, email

Description automatically generated

**Figure 3**

*Dialog used to Participate in a Group, Join a Group, or Manage Permissions*

Graphical user interface, text, application

Description automatically generated Graphical user interface, text, application, chat or text message

Description automatically generated

**Figure 4**

*CC Comments and Tags for the TBIBank Grand Rounds Exercise*

Graphical user interface, text, application

Description automatically generated

# References

Bernstein-Ratner, N., & MacWhinney, B. (2018). Fluency Bank: A new resource for fluency research and practice. *Journal of Fluency Disorders*, *56*, 69-80. <https://doi.org/10.1016/j.jfludis.2018.03.002>

Boullosa, B., de Castilho, R. E., Laskari, N. K., Klie, J.-C., & Gurevych, I. (2018). Integrating knowledge-supported search into the inception annotation platform. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations,

Brouwer, M., Brugman, H., & Kemps-Snijders, M. (2017). MTAS: A Solr/Lucene based multi tier annotation search solution. CLARIN 2016, Aix-en-Provence.

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, *31*(10), 1105-1126.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, *26*(10), 1250-1277. <https://doi.org/10.1080/02687038.2012.693584>

De Castilho, R. E., Mújdricza-Maydt, E., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., & Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures. Proceedings of the workshop on language technology resources and tools for digital humanities (LT4DH), Osaka, Japan.

Farzana, S., & Parde, N. (2022). Are Interaction Patterns Helpful for Task-Agnostic Dementia Detection? An Empirical Exploration. Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue,

Farzana, S., Valizadeh, M., & Parde, N. (2020). Modeling dialogue in conversational cognitive health screening interviews. Proceedings of the 12th Language Resources and Evaluation Conference,

Forbes, M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: a resource for clinicians. *Seminars in Speech and Language*, *33*(3), 217-222. <https://doi.org/10.1055/s-0032-1320041>

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons [Article]. *Applied Linguistics*, *21*(3), 354-375. <http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=44400914&site=ehost-live>

Girardi, C., Speranza, M., Sprugnoli, R., & Tonelli, S. (2014). Cromer: A tool for cross-document event and entity coreference. Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland.

Haider, F., De La Fuente, S., & Luz, S. (2019). An assessment of paralinguistic acoustic features for detection of Alzheimer's Dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 272-281.

Holland, A., Fromm, D., Forbes, M., & MacWhinney, B. (2014). *The famous people protocol* International Aphasia Rehabilitation Conference, The Hague.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test (2nd ed.)*. Lippincott Williams & Wilkins.

Kertesz, A. (2007). *Western Aphasia Battery (Revised)*. PsychCorp, San Antonio, Tx.

Lee, L. (1966). Developmental sentence types: A method for comparing normal and deviant syntactic development. *Journal of Speech and Hearing Disorders*, *31*, 331-330.

Luz, S., Haider, F., de la Fuente Garcia, S., Fromm, D., & MacWhinney, B. (2021). Editorial: Alzheimer’s dementia recognition through spontaneous speech. *Frontiers in Aging Neuroscience*, *3*. <https://doi.org/10.3389/fcomp.2021.780169>

MacWhinney, B., Martell, C., Schmidt, T., Wagner, J., Wittenburg, P., Brugman, H., Broeder, D., & Hoffert, E. (2004). Collaborative commentary: Opening up spoken language databases. In *LREC 2004* (pp. 11-15). LREC. <https://psyling.talkbank.org/years/2004/LREC-cc.pdf>

Mann, W. C., & Thompson, S. A. (1992). *Discourse description*. John Benjamins.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021. <https://doi.org/10.1038/s41562-016-0021>

O'Reilly, C., Iavarone, E., & Hill, S. L. (2017). A framework for collaborative curation of neuroscientific literature. *Frontiers in neuroinformatics*, *11*, 1-16.

Overton, S., & Wren, Y. (2014). Outcome measurement using naturalistic language samples: A feasibility pilot study using language transcription software and speech and language therapy assistants. *Child Language Teaching and Therapy*, *30*, 221-229.

Papazian, F., Bossy, R., & Nédellec, C. (2012). AlvisAE: a collaborative Web text annotation editor for knowledge acquisition. Sixth Linguistic Annotation Workshop,

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): A computerized text analysis program*. Lawrence Erlbaum Associates.

Rochon, E., Saffran, E., Berndt, R., & Schwartz, M. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, *72*, 193-218. <https://doi.org/10.1006/brln.1999.2285>

Scarborough, H. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, *11*(1), 1-22. <https://doi.org/10.1017/S0142716400008262>

Tablan, V., Bontcheva, K., Roberts, I., & Cunningham, H. (2015). Mímir: An open-source semantic search framework for interactive information seeking and discovery. *Web Semantics: Science, Service, and Agents on the World Wide Web*, *30*, 52-68.

Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L., & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language*, *51*, 124-129.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS ONE*, *13*(2), e0192773. <https://doi.org/10.1371/journal.pone.0199469>

**Figure 1:**

*Starting Page for Information about CC --* [*https://talkbank.org/CC/*](https://talkbank.org/CC/)

Text, timeline

Description automatically generated with medium confidence

**Figure 2**

*Starting page for the TalkBank Browser with the Collab Button*

Graphical user interface, text, application, email

Description automatically generated

**Figure 3**

*Dialog used to Participate in a Group, Join a Group, or Manage Permissions*

Graphical user interface, text, application

Description automatically generated Graphical user interface, text, application, chat or text message

Description automatically generated

**Figure 4**

*CC Comments and Tags for the TBIBank Grand Rounds Exercise*

Graphical user interface, text, application

Description automatically generated