Group 1: Features of Dataset Applications

User Interface Features Typical of Dataset Applications

The code for each dsC data set includes a customized "Dataset Application" which displays individual samples and groups of samples via 2D, 3D, and native-compiled GUI controls. Each Dataset Application can thereby make use of advanced visual and interactive features that are uniquely possible when using customized, native-compiled GUI classes. The following screenshots will show several examples of these features, including:

Specialized Top-Level Controls Tree Widgets, Stacked Widgets, and Graphics Scenes.

Context Menus Systematically organize functionality around UI layouts.

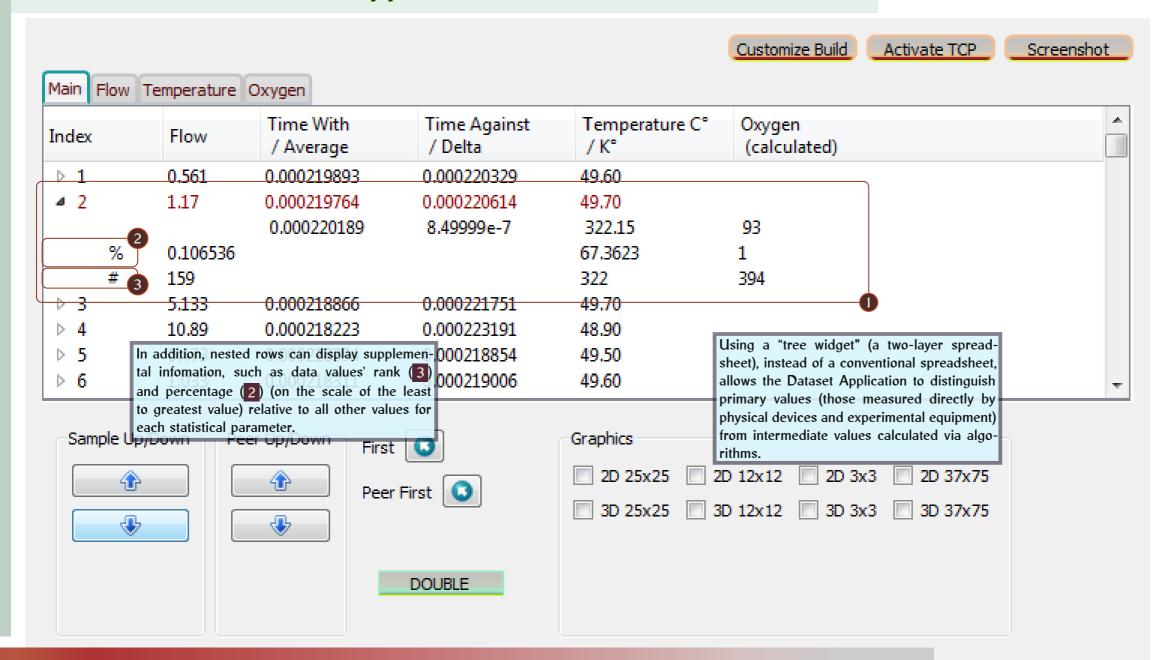
Multi-Window Displays Divide application functionality in multiple specialized top-level windows and/or dialog boxes.



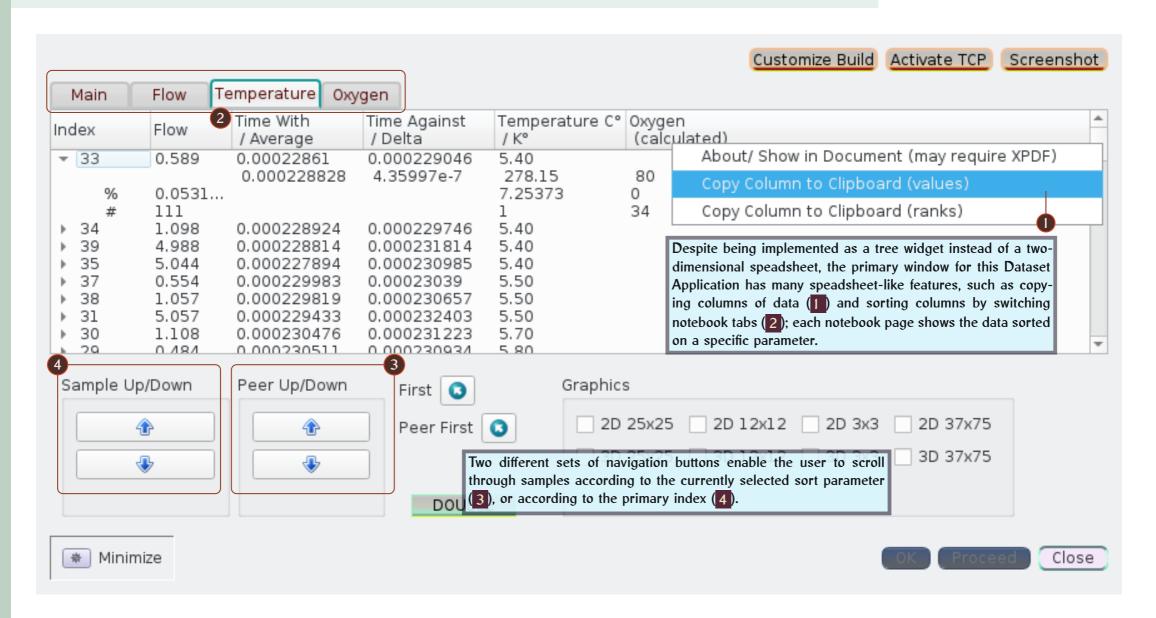
Linguistic Technology Systems



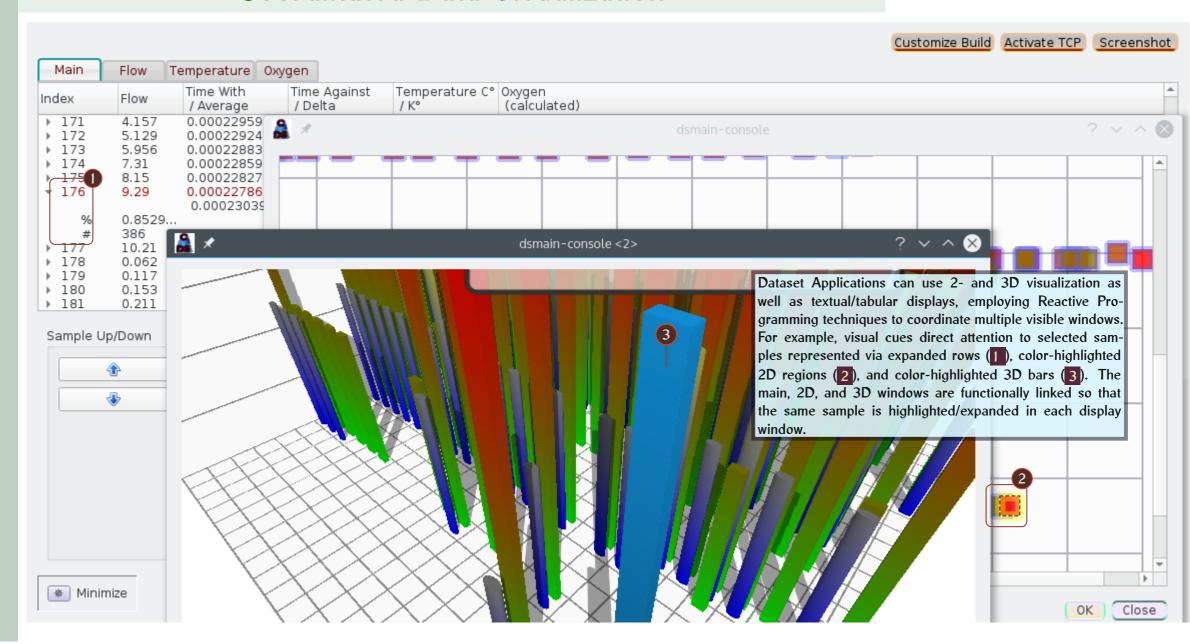
Initial Application Window



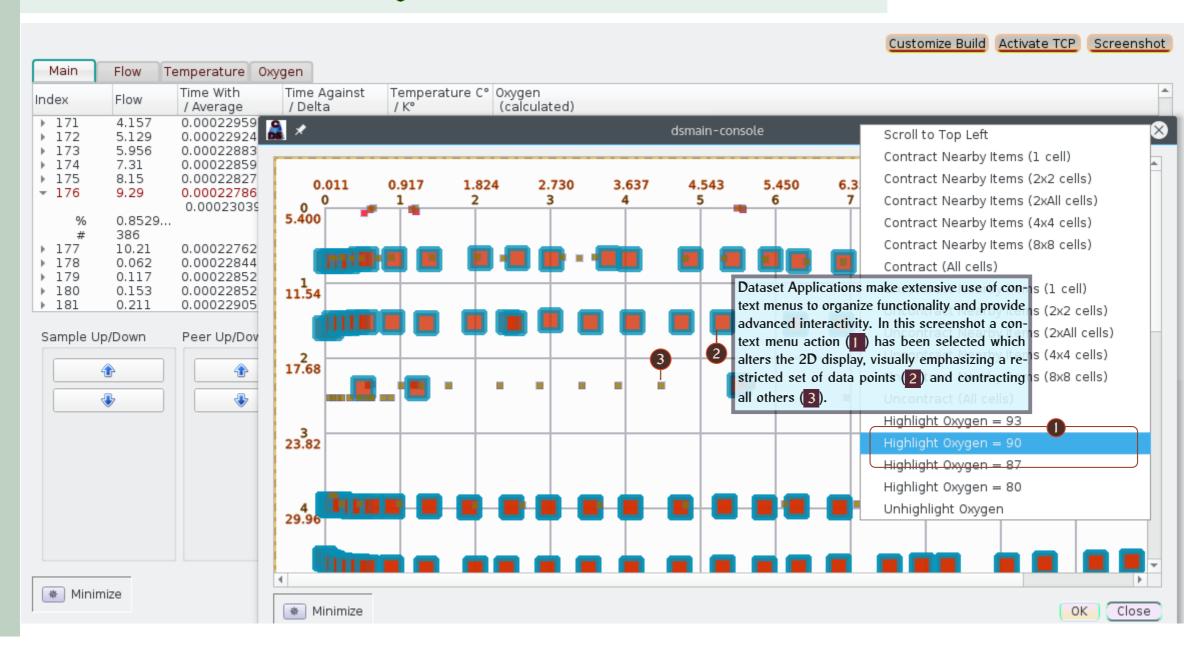
Interacting with the Main Window



Coordinated Data Visualization



Interacting with the Visuals



Getting Information About Modeling Parameters

Using Dataset Applications as Pedagogical Tools

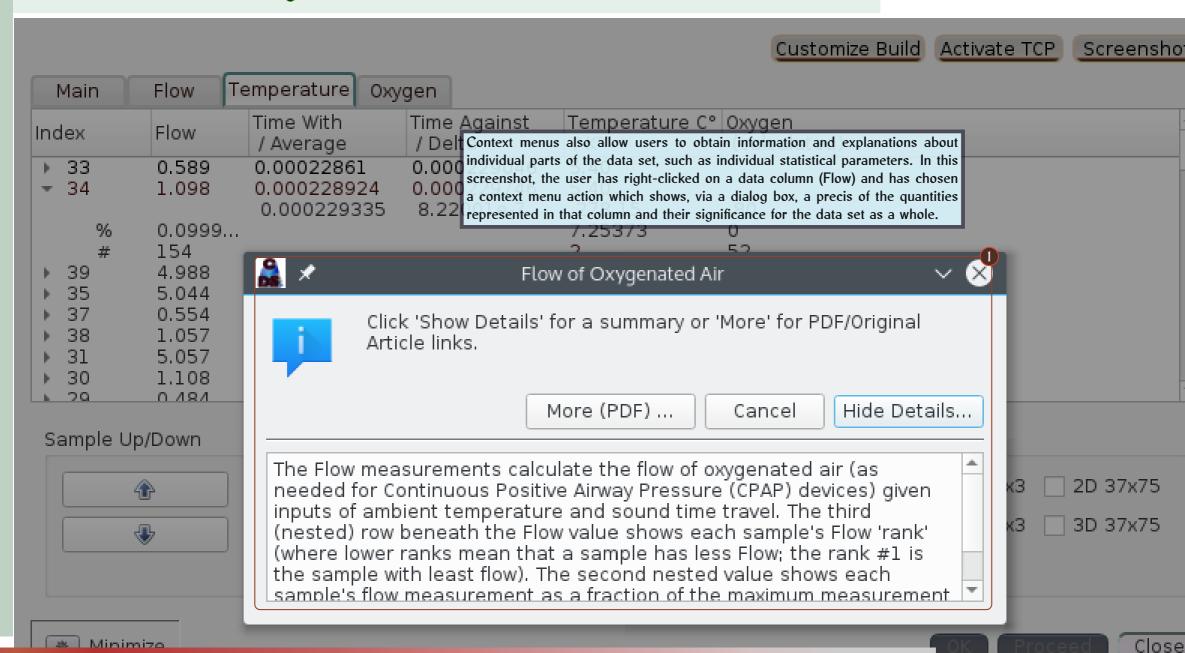
In addition to interactive visualization, Dataset Applications are useful tools for understanding experimental protocols and research methods. Within Dataset Applications, modeling units such as statistical parameters and record fields are visible in situ within a GUI — identified by labels, buttons, and other interactive micro-controls. As a result, users encounter modeling elements in a structured visual-intractive context. To learn more about modeling elements, Dataset Applications are equipped with several pedagogical features shown on the following screenshots:

"About" Dialogs Brief summaries of research trms and parameters.

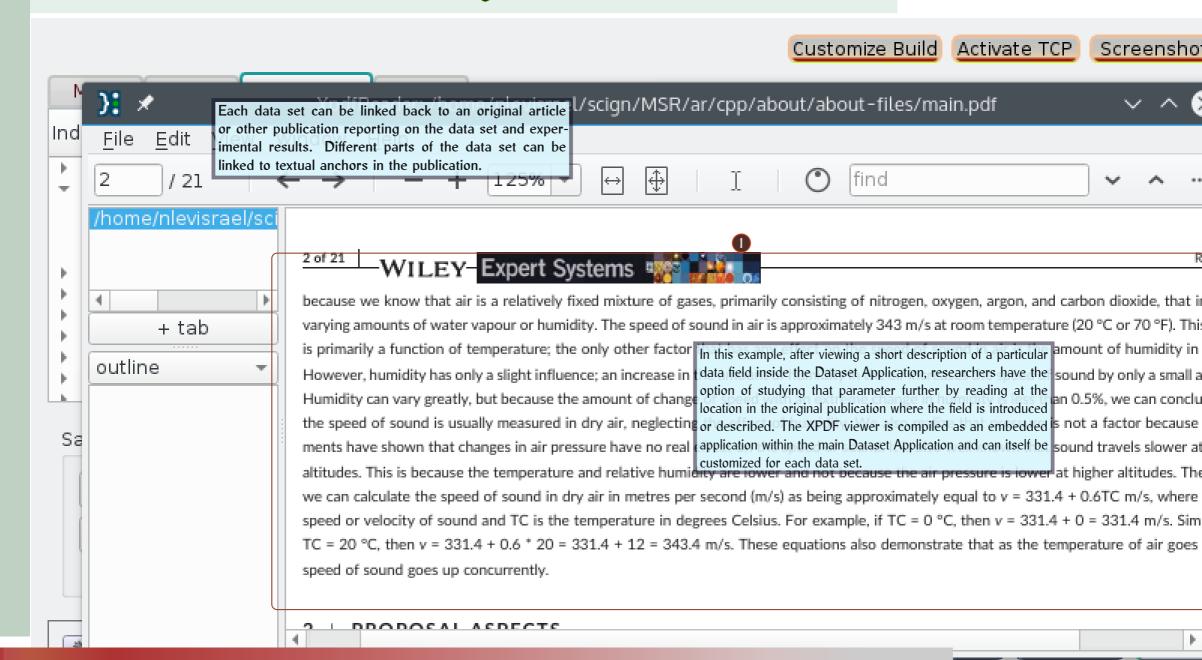
XPDF Links Links back to research articles read in an embedded PDF viewer.

XPDF Enhancements The XPDF viewer can be customized for each data set and included with dataset code, with extra features to integrate article or book texts with Dataset Applications.

Obtaining Information About Parameters



Embedding XPDF



Testing and Fine-Tuning Dataset Applicationss

Tools for Editors and Developers

Although ordinary users can explore and visualize dsC data sets "Out of the Box", advanced users have many options for customizing their build of the application in terms of their specific roles and available 3rd-party libraries. These fine-tuning possibilities include:

Test Suites Tools for creating and/or running test suites to ensure that the Dataset Application works across platforms.

Data Export Tools for reusing data in other projects.

External Libraries Some features like XPDF and 3D graphics require libraries that cannot be published with the data set in source code form. Advanced users can slect which of these libraries to incoporate into thir version of the Dataset Application.

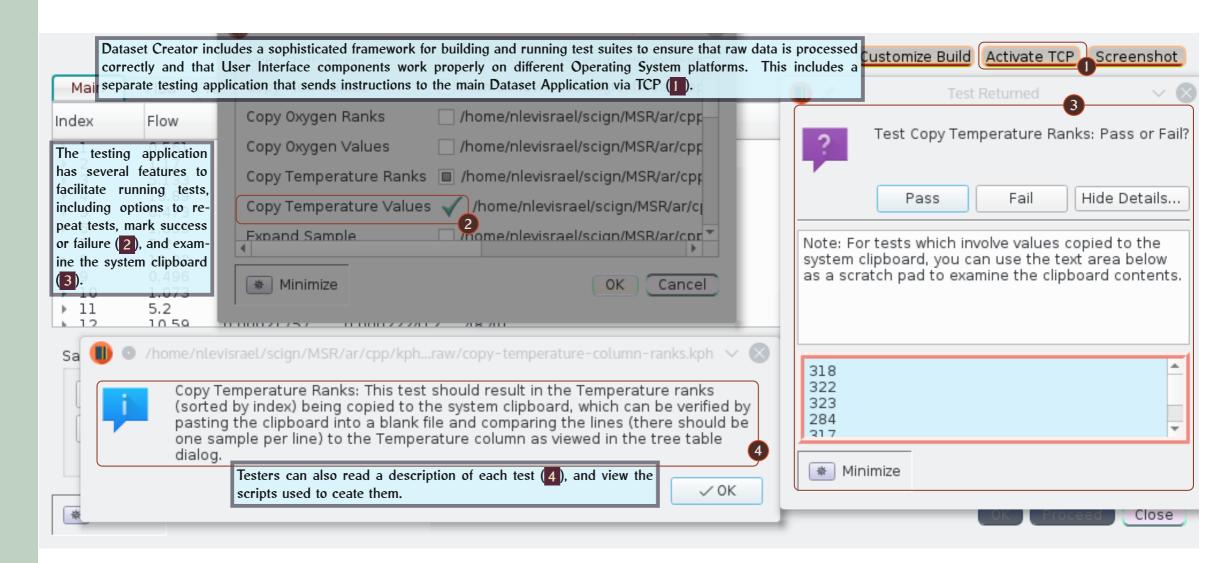
Scripting Data sets can comoile their own scripting environment to automate testing and manipulation of research data.

Networking Dataset Applications can use an embedded TCP server to communicate with other applications, enabling multi-application workflows (this is also how testing is implemented).

Configuring the Data Set Application

	Operating System Profile	Customize Build Activate TCP Screenshot
n * * * * * * *	Linux (Generic) ▼ 32 Bit ▼ 64 Bit	Oxygen
	Compile Options Use 3d graphics Use Kauvir/Phaon and TCP (for tests) Qt PNG/FreeType libraries V Use XPDF System PNG/FreeType libraries Build KDMI Components and Console (for data export)	Calc Using Qt Creator, the Dataset Creator will automatically launch the main Dataset Application with every feature needed in order to visualize and explore the data. In addition, the data set includes several configurations allowing users to incorporate more specialized or complex features, such as XPDF, test suites, and data export code. Users can fine-tune which additional features they wish to utilize — via a separate dialog box [1] and [2] — to create a customized build of the main Dataset Application and supplemental executables.
	▼ Build Research Object Information Console	
L	▼ Build External XPDF Application	
9	Preview (right click "Administrator" to enable/disable)	
	Reset (reset files to original state; right-click "Administrator" to enable/disable) Select User Role	25 2D 12x12 2D 3x3 2D 37x75 The Dataset Creator also recognizes distinct "roles" (2), including general readers, authors, those who double-check the main
	User, Reader, Researcher (Default) ✓ Author Editor Tester Administrator	Dataset Application via a test suite, and those who design the test suite and write dataset code overall (dubbed "Administrators").

Testing the Data Set Application



Features of Dataset Applications for Books

Datasets Compiled From Book Examples

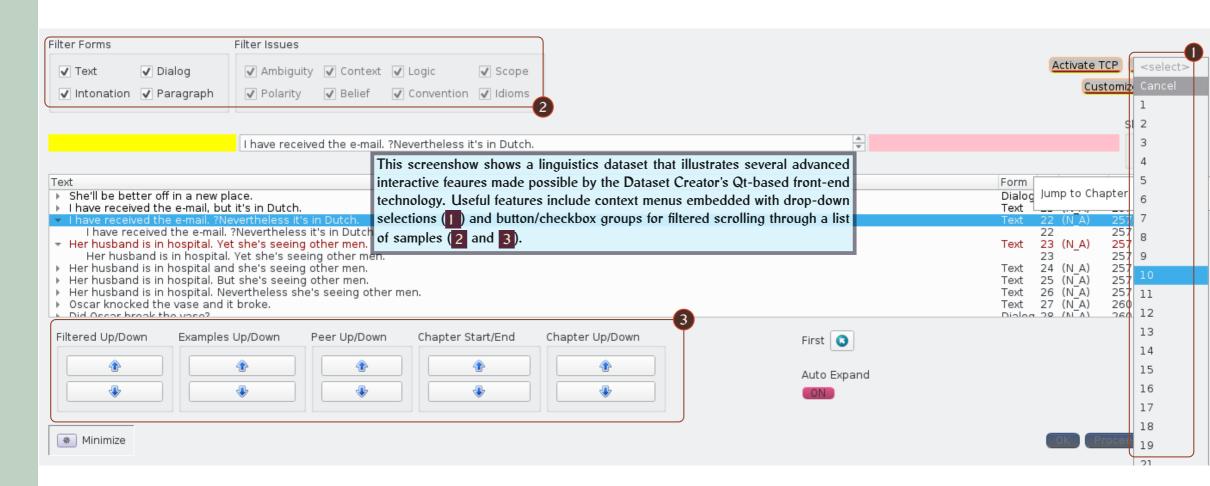
The remaining screenshots demonstrate how data sets can be used even outside of a lab contxt generating experiment data. The pictured data set represents a corpus of linguistic examples mined from Wiley's *Blackwell Handbook of Pragmatics*. Creating data sets from book-length publications can encompass several steps:

Text Mining In the case of linguistics, this involves locating example sentences within linguistics texts and storing them as an independent corpus.

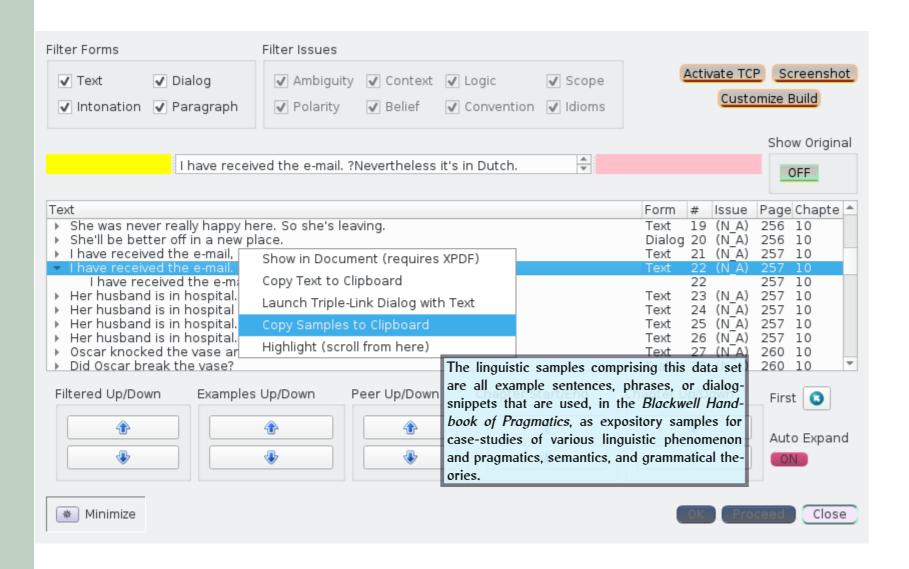
Canonical Formatting If possible, linguistics texts should be annotated so that extracting exmples can be automated.

Annotation Linguistic corpuses are often annotated to identify structural details, beyond raw text, in each sample.

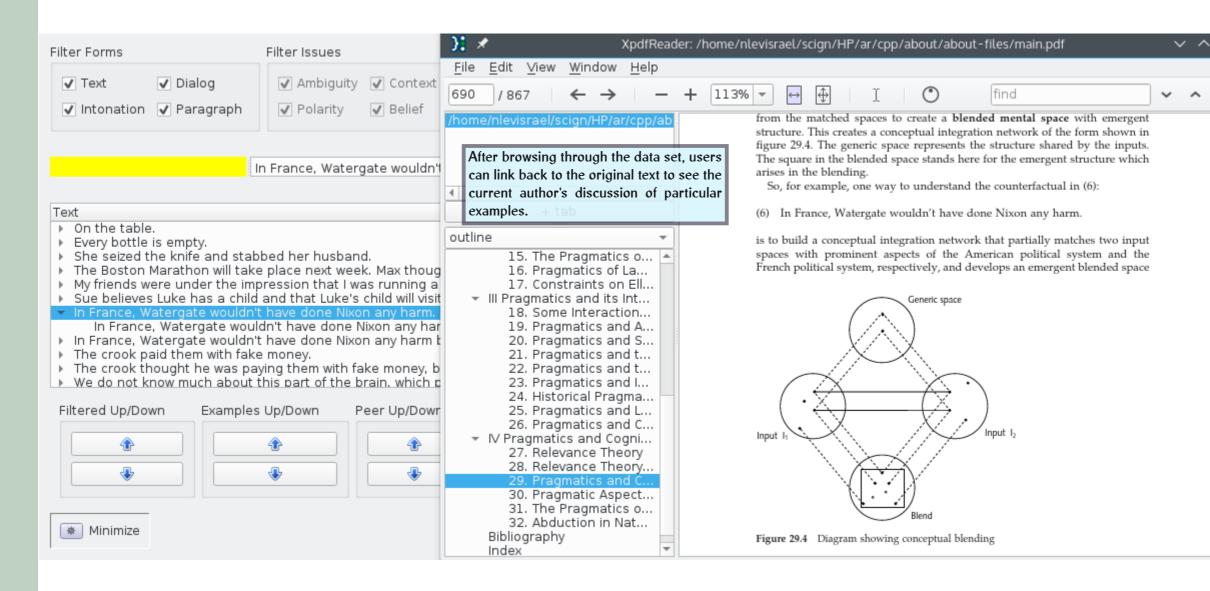
Creating a Data Set from a Book



Interacting with Data Samples



Linking Back to the Book



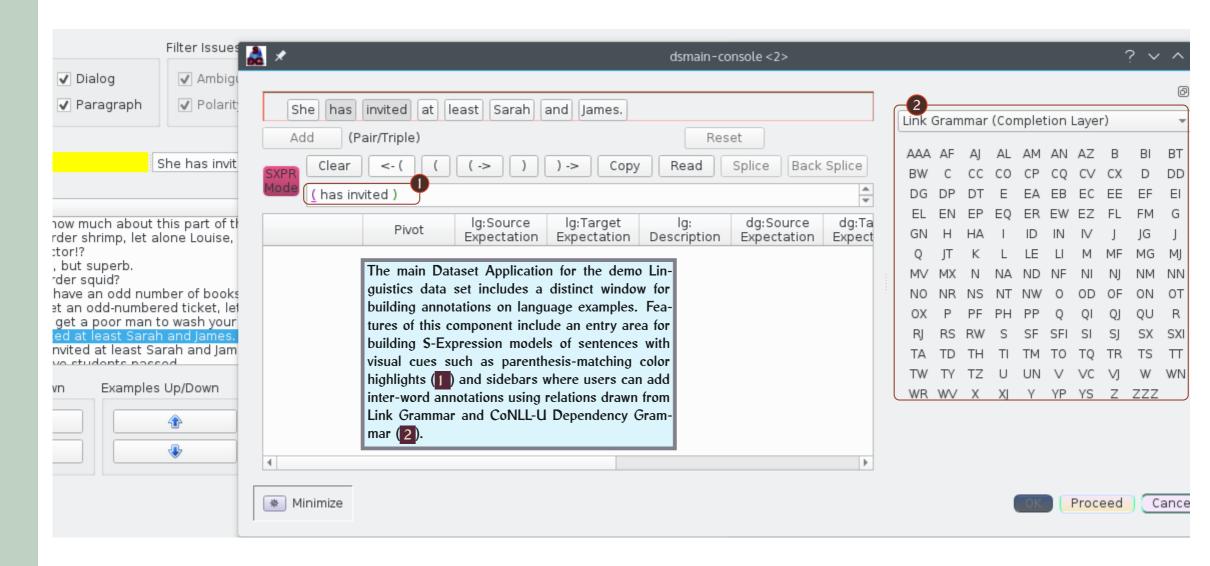
A Linguistics Annotation System

Tools to Facilitate Annotating Linguistic Corpora

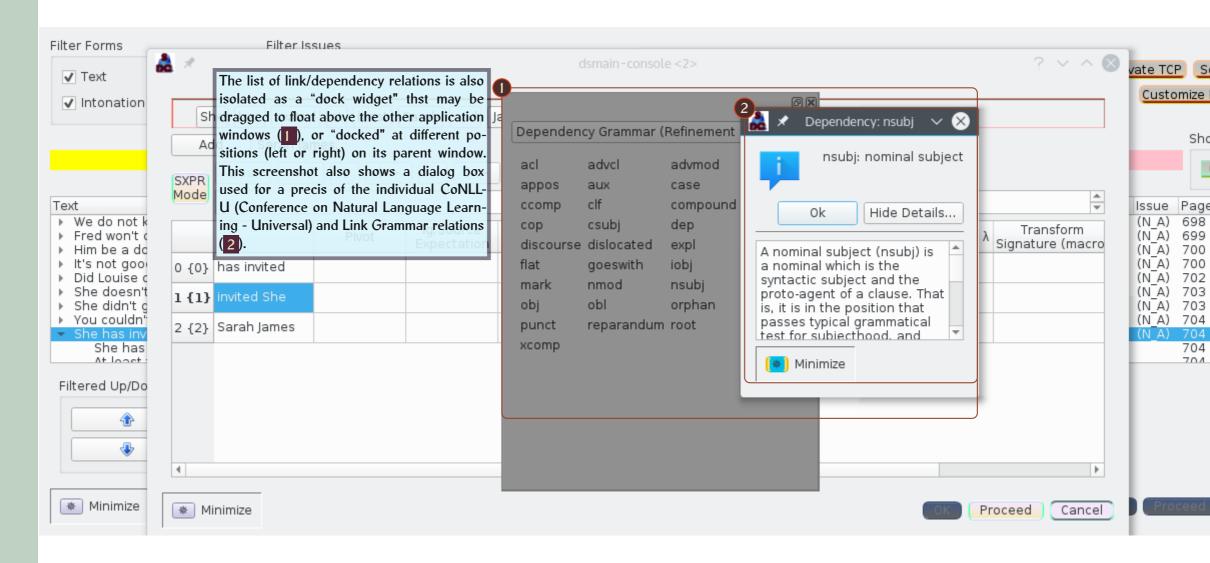
The final three screenshots show an example of how a custom-signd application can facilotat the task of building an annotated corpus from a linguistics text. The components demonstrated here enable several strategies (which can be combined) for dscribing parsing structures and the logical composition of language samples:

- S-Expressions Representing linguistic units as semantic and syntactic transformations triggered by words assigned to "functional" types.
- Deepndency Grammar Representing phrase structures viabinter-word syntactic relationships.
- Link Grammar Representing linguistic structure via connectors internal to each word-sense. Inter-word links are activated when each word in the pair has a connector compatible with the other word's connector. Intuitively, a connector represents how one word's meaning or grammatic contribution can be "completed" by linking to a separate word.

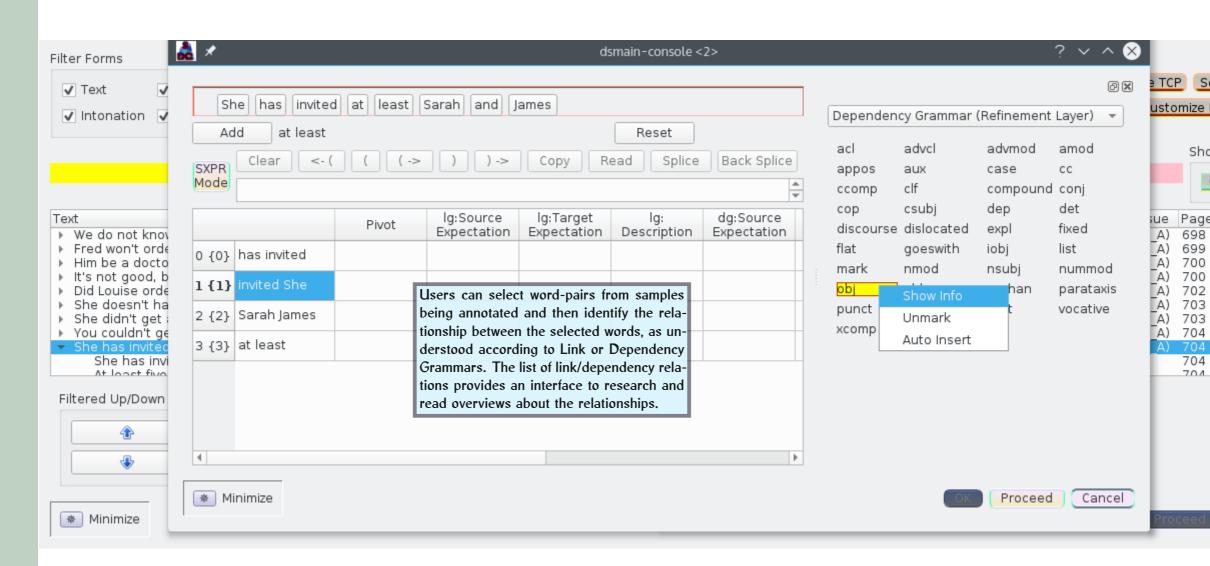
Building Parsing Models



Using Dock Widgets For Flexible Layout

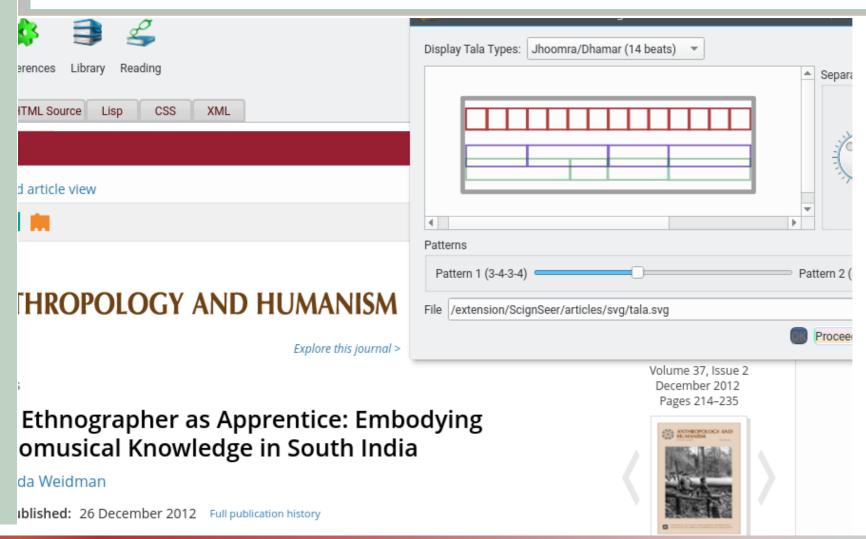


Link and Dependency Grammar Annotations



A3R Document Viewers

A3R applications may embed viewers for document formats such as e-Pub, HTML, and PDF; then supplement conventional publications with special components customized for individual manuscripts: e.g. (as in this case), a widget allowing readers to visually explore patterns in classical Indian music.



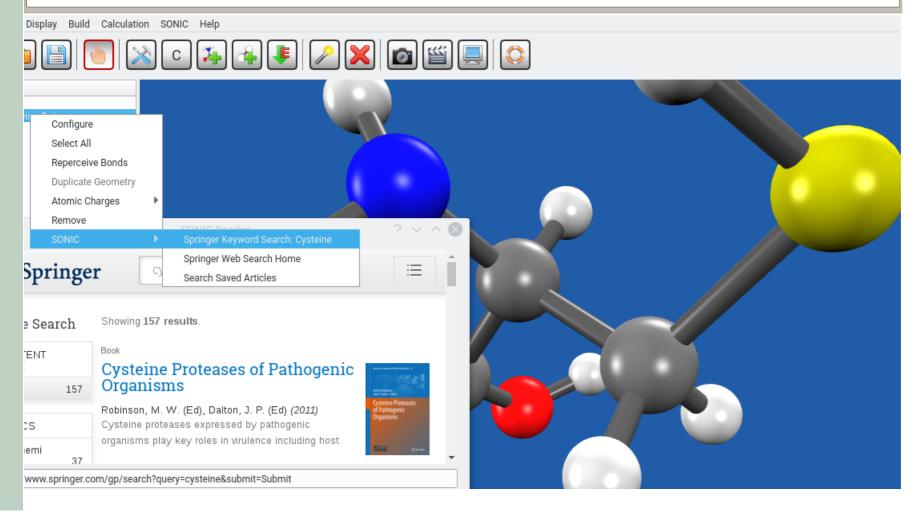


Linguistic Technology Systems



A3R Document Viewers as Embedded Components

Document Viewers may also be embedded in host applications which provide domain-specific visualization capabilities. For example, chemistry papers might be viewed within IQmol (a Qt-based program for molecular visualization and physical/chemical analysis) via an A3R document-viewer plugin.





Linguistic Technology Systems

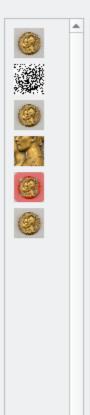


Document Viewers Augmented With APIs

Another strategy for interactive publications is linking documents with APIs maintained by publishers, or by cultural or educational institutions.



API Open Folder



Instructions

View

As an example, documents mentioning artifacts held in a museum can provide features to view more information about those museum-pieces through the host institution's API.



MEDAL

Click the icon to save this object

This is a Medal. We acquired it in 1920. Its moving is a part of the Product Design and Decorative department.

Cite this object as

Medal; bronze; 1920-31-1

Embedded Multimedia

Custom-built A3R document viewers can provide convenient access to multimedia content embedded in or linked to texts — including audio files, videos, and 3D graphics scenes or models.

In this case a video player is launched in a dialog box, floating above the article text. For those reading digital books or articles, videos and other multimedia content can be presented through secondary windows launched via context menus; text and multimedia may thereby be viewed side-by-side.



Behavior

The

ab

Red pandas are generally solitary, but there are a couple of develop extended associations with their mothers that las breeding season.



In terms of t tend to have other. This n search for th

>rk.org/red_panda/about-the-red-panda/





Linguistic Technology Systems

