This screenshot shows a traditional application (not using dsC) displaying the current data set as originally presented in a spreadsheet.

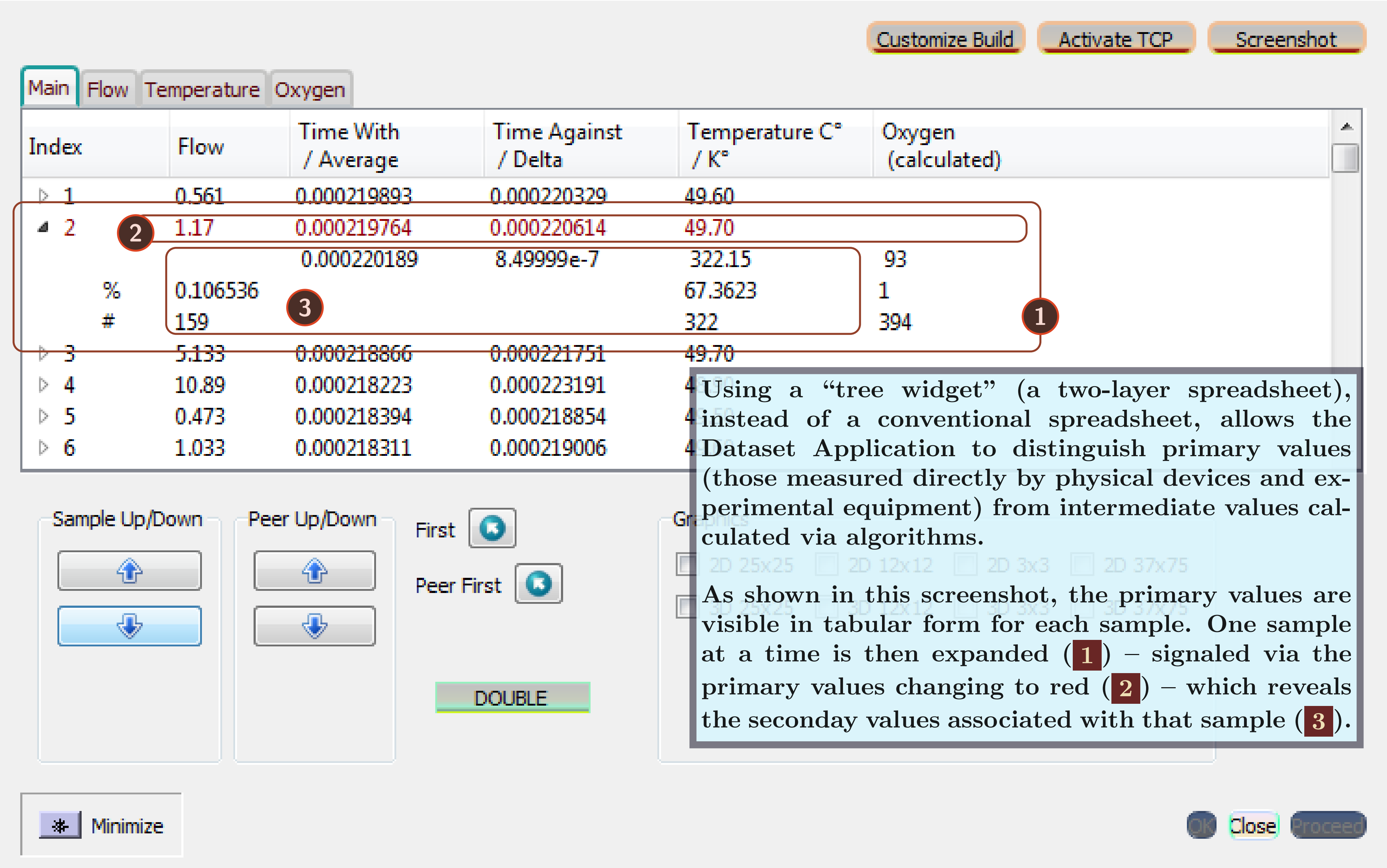
We include this example to show several lacunae in spreadsheet applications' functionality that can be improved by using a customized Dataset Application. The following screenshots will highlight three limitations in particular:

- The spreadsheet does not store information in a format amenable to reuse by other projects or by production deployments of the author's method. By contrast, the dsC version models all data as standalone, cross-platform C++ classes that can be reused for any new project.
- The spreadsheet does not have an obvious mechanism for researching individual modeling elements. For instance, there is no explanation near the column labels "WithFlow" or "Against" which explain what these parameters mean and how they are used.
- The spreadsheet will not be suitable for large-scale or commercial deployment. For example, the touchscreen interface to a device monitoring airflow needs a UI specific to its cyberphysical data — it is not feasible for medical devices to run spreadsheet software!
- The spreadsheet groups all statistical parameters (i.e., column headers) together, without representing their internal organization. For example, the spreadsheet does not structurally distinguish between raw cyberphysical inputs, intermediate calculated values, and the important computed values which the cyberphysical system is designed to produce. In this case, as the original article explains, Oxygenated air flow and Oxygen concentration levels are the significant derived values, and the research presents a method for computing these values given a specific cyberphysical and algorithmic setup.

	B	C	D	E	F	G	H	I
4	Concentration		Time (Seconds)					
5	%O2	Flow (Lpm)	WithFlow	Against	Temperature (°C)	avgTime	Delta Time	Temperature (°K)
6	93	0.561	0.000219892800	0.000220328700	49.60	0.000220110750	0.000000435900	322.7499
7	93	1.170	0.000219764300	0.000220614400	49.70	0.000220189350	0.000000850100	322.85
8	93	5.133	0.000218866400	0.000221751100	49.70	0.000220308750	0.000002884700	322.85
9	93	10.890	0.000218222600	0.000223191400	48.90	0.000220707000	0.000004968800	322.05
10	80	0.473	0.000218394100	0.000218854200	49.50	0.000218624150	0.000000460100	322.65
11	80	1.033	0.000218310700	0.000219005600	49.60	0.000218658150	0.000000694900	322.7499
12	80	5.200	0.000217227400	0.000220173500	49.70	0.000218700450	0.000002946100	322.85
13	80	10.220	0.000216661100	0.000221369300	48.90	0.000219015200	0.000004708200	322.05

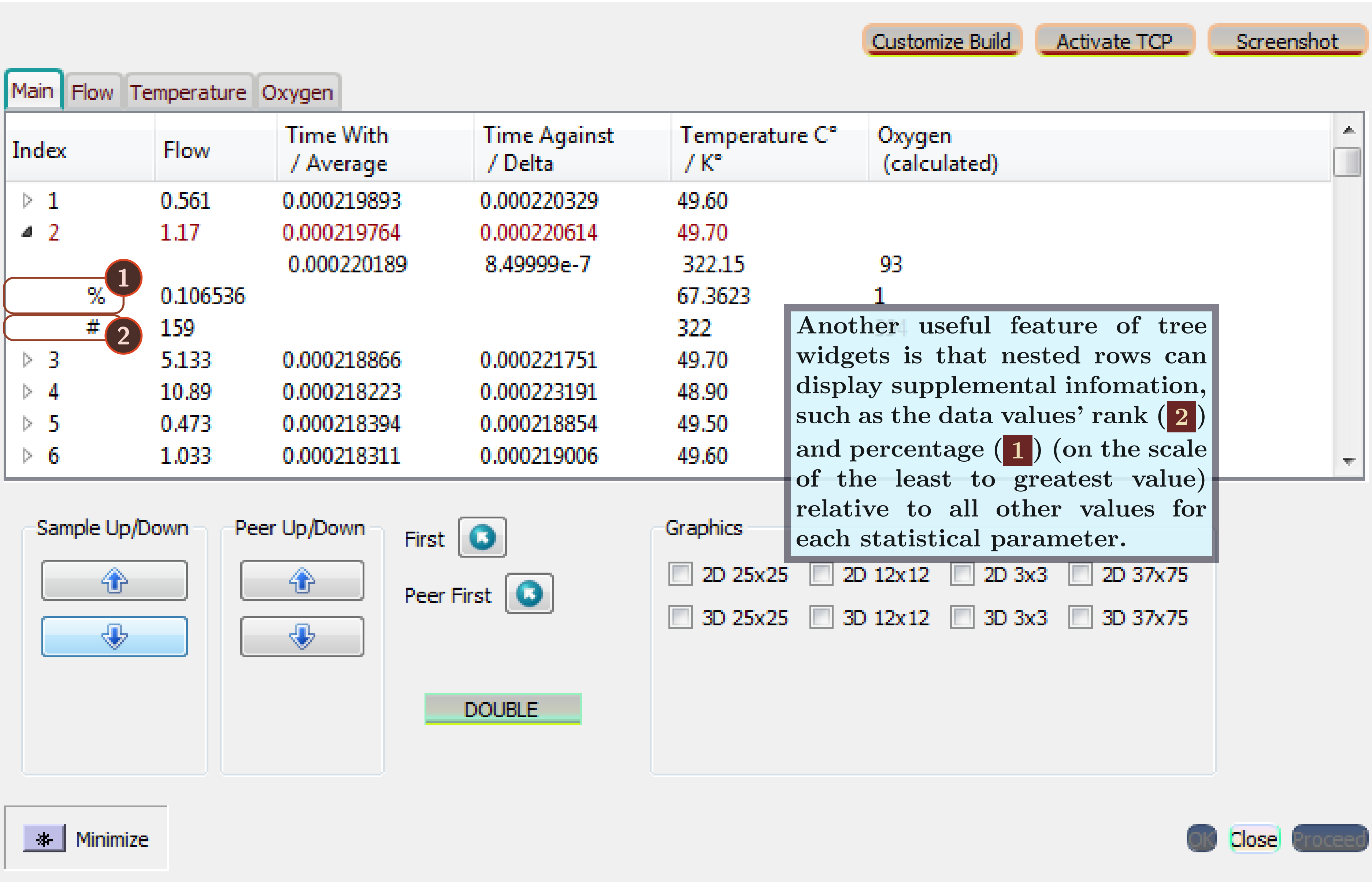
System Overview | Data

Sheet 2 of 2 | 1 rows, 3 columns selected | PageStyle_Data | Average: ; Sum: 0 | 100%

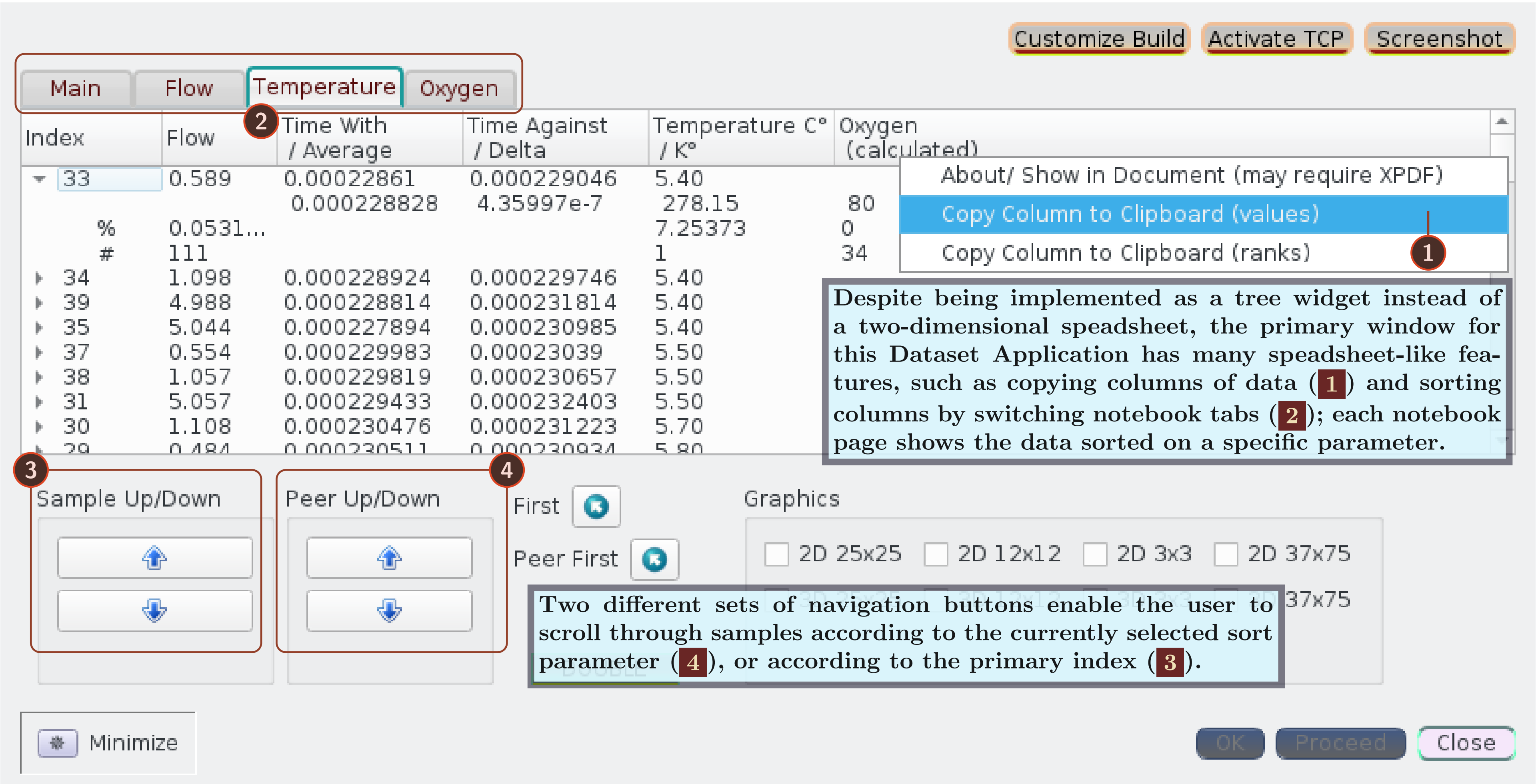


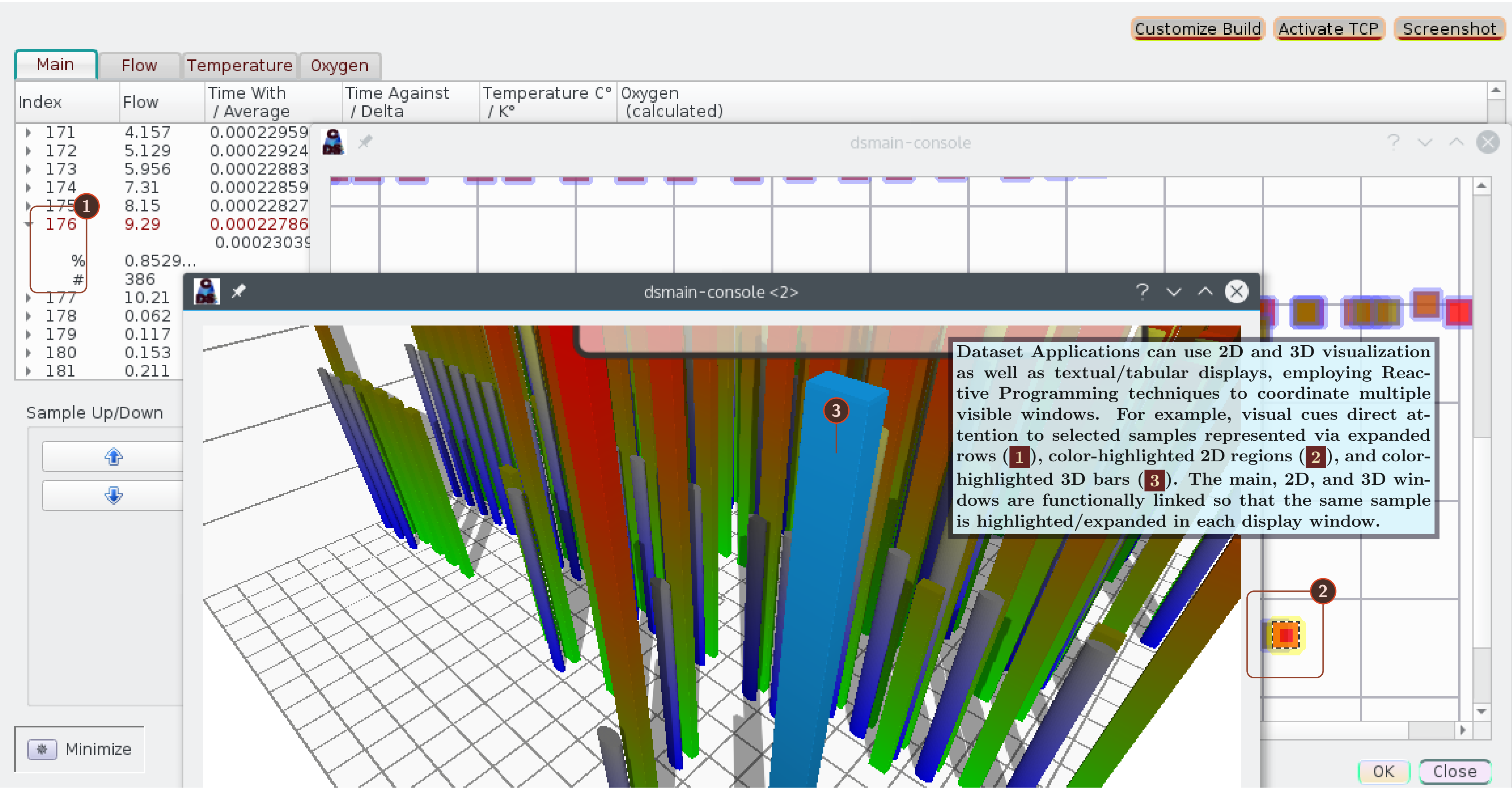
Using a "tree widget" (a two-layer spreadsheet), instead of a conventional spreadsheet, allows the Dataset Application to distinguish primary values (those measured directly by physical devices and experimental equipment) from intermediate values calculated via algorithms.

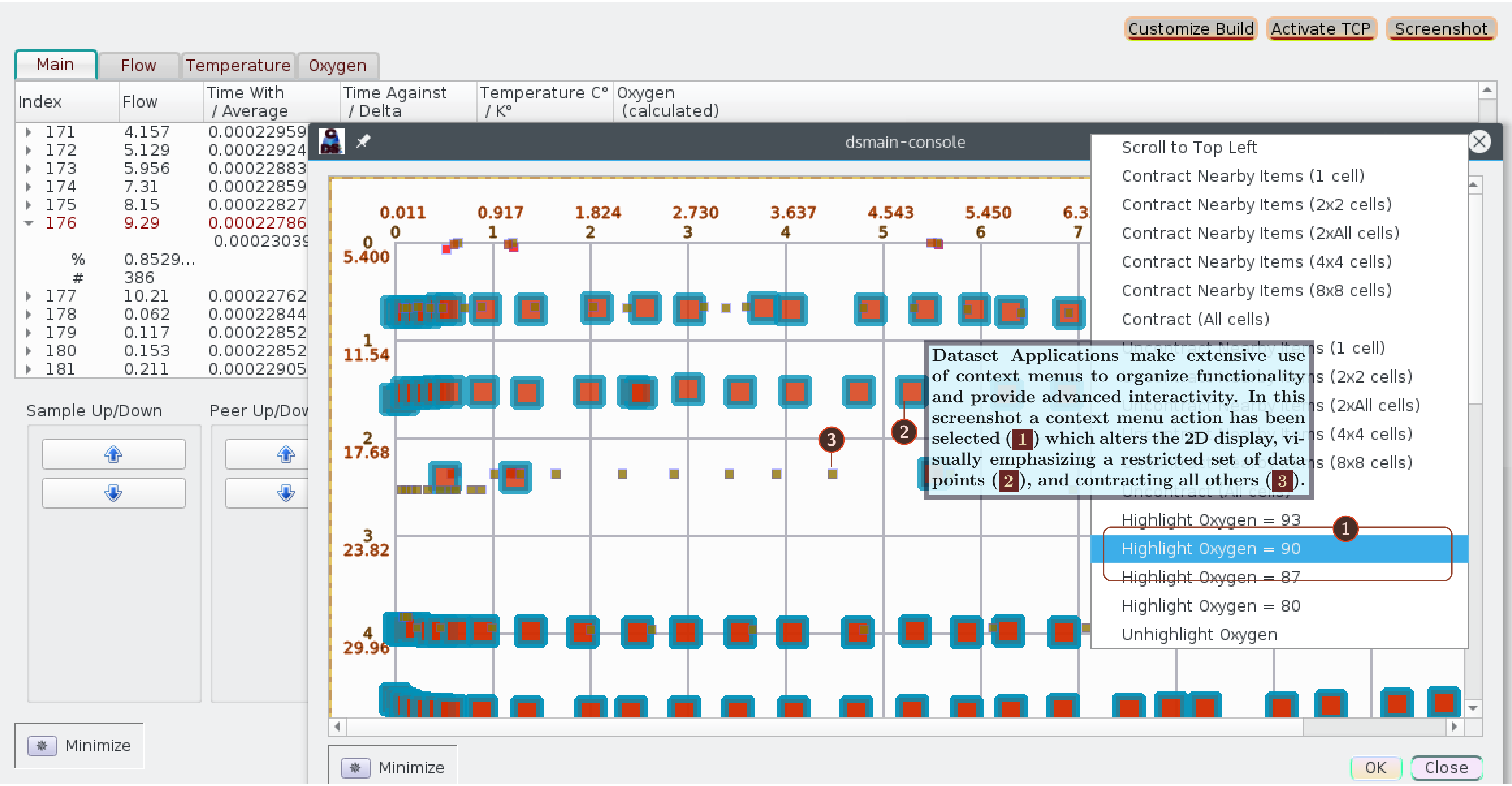
As shown in this screenshot, the primary values are visible in tabular form for each sample. One sample at a time is then expanded (1) – signaled via the primary values changing to red (2) – which reveals the seconday values associated with that sample (3).



Another useful feature of tree widgets is that nested rows can display supplemental infomation, such as the data values' rank (2) and percentage (1) (on the scale of the least to greatest value) relative to all other values for each statistical parameter.


Customize Build
Activate TCP
Screenshot
Main
Flow
Temperature
Oxygen
Index
Flow
Time With / Average
Time Against / Delta
Temperature C° / K°
Oxygen (calculated)
About/ Show in Document (may require XPDF)
Copy Column to Clipboard (values)
Copy Column to Clipboard (ranks)
Despite being implemented as a tree widget instead of a two-dimensional speadsheet, the primary window for this Dataset Application has many speadsheet-like features, such as copying columns of data (1) and sorting columns by switching notebook tabs (2); each notebook page shows the data sorted on a specific parameter.
Sample Up/Down
Peer Up/Down
First
Peer First
Graphics
2D 25x25
2D 12x12
2D 3x3
2D 37x75
Two different sets of navigation buttons enable the user to scroll through samples according to the currently selected sort parameter (4), or according to the primary index (3).
Minimize
OK
Proceed
Close


Customize Build
Activate TCP
Screenshot
Main
Flow
Temperature
Oxygen
Index
Flow
Time With / Average
Time Against / Delta
Temperature C° / K°
Oxygen (calculated)
171 4.157 0.00022959
172 5.129 0.00022924
173 5.956 0.00022883
174 7.31 0.00022859
175 8.15 0.00022827
176 9.29 0.00022786
0.00023039
% 0.8529...
# 386
177 10.21
178 0.062
179 0.117
180 0.153
181 0.211
Sample Up/Down
Minimize
dsmain-console
dsmain-console <2>
1
2
3
Dataset Applications can use 2D and 3D visualization as well as textual/tabular displays, employing Reactive Programming techniques to coordinate multiple visible windows. For example, visual cues direct attention to selected samples represented via expanded rows (1), color-highlighted 2D regions (2), and color-highlighted 3D bars (3). The main, 2D, and 3D windows are functionally linked so that the same sample is highlighted/expanded in each display window.
OK
Close


Customize Build
Activate TCP
Screenshot
Main
Flow
Temperature
Oxygen
Index
Flow
Time With / Average
Time Against / Delta
Temperature C° / K°
Oxygen (calculated)
171 4.157 0.00022959
172 5.129 0.00022924
173 5.956 0.00022883
174 7.31 0.00022859
175 8.15 0.00022827
176 9.29 0.00022786
0.00023039
% 0.8529...
# 386
177 10.21 0.00022762
178 0.062 0.00022844
179 0.117 0.00022852
180 0.153 0.00022852
181 0.211 0.00022905
Sample Up/Down
Peer Up/Dov
Minimize
dsmain-console
0.011 0.917 1.824 2.730 3.637 4.543 5.450 6.3
0 1 2 3 4 5 6 7
0 5.400
1 11.54
2 17.68
3 23.82
4 29.96
Scroll to Top Left
Contract Nearby Items (1 cell)
Contract Nearby Items (2x2 cells)
Contract Nearby Items (2xAll cells)
Contract Nearby Items (4x4 cells)
Contract Nearby Items (8x8 cells)
Contract (All cells)
s (1 cell)
s (2x2 cells)
s (2xAll cells)
s (4x4 cells)
s (8x8 cells)
Highlight Oxygen = 93
Highlight Oxygen = 90
Highlight Oxygen = 87
Highlight Oxygen = 80
Unhighlight Oxygen
Dataset Applications make extensive use of context menus to organize functionality and provide advanced interactivity. In this screenshot a context menu action has been selected (1) which alters the 2D display, visually emphasizing a restricted set of data points (2), and contracting all others (3).
1
2
3
OK
Close

Getting Information About Modeling Parameters

Using Dataset Applications as Pedagogical Tools

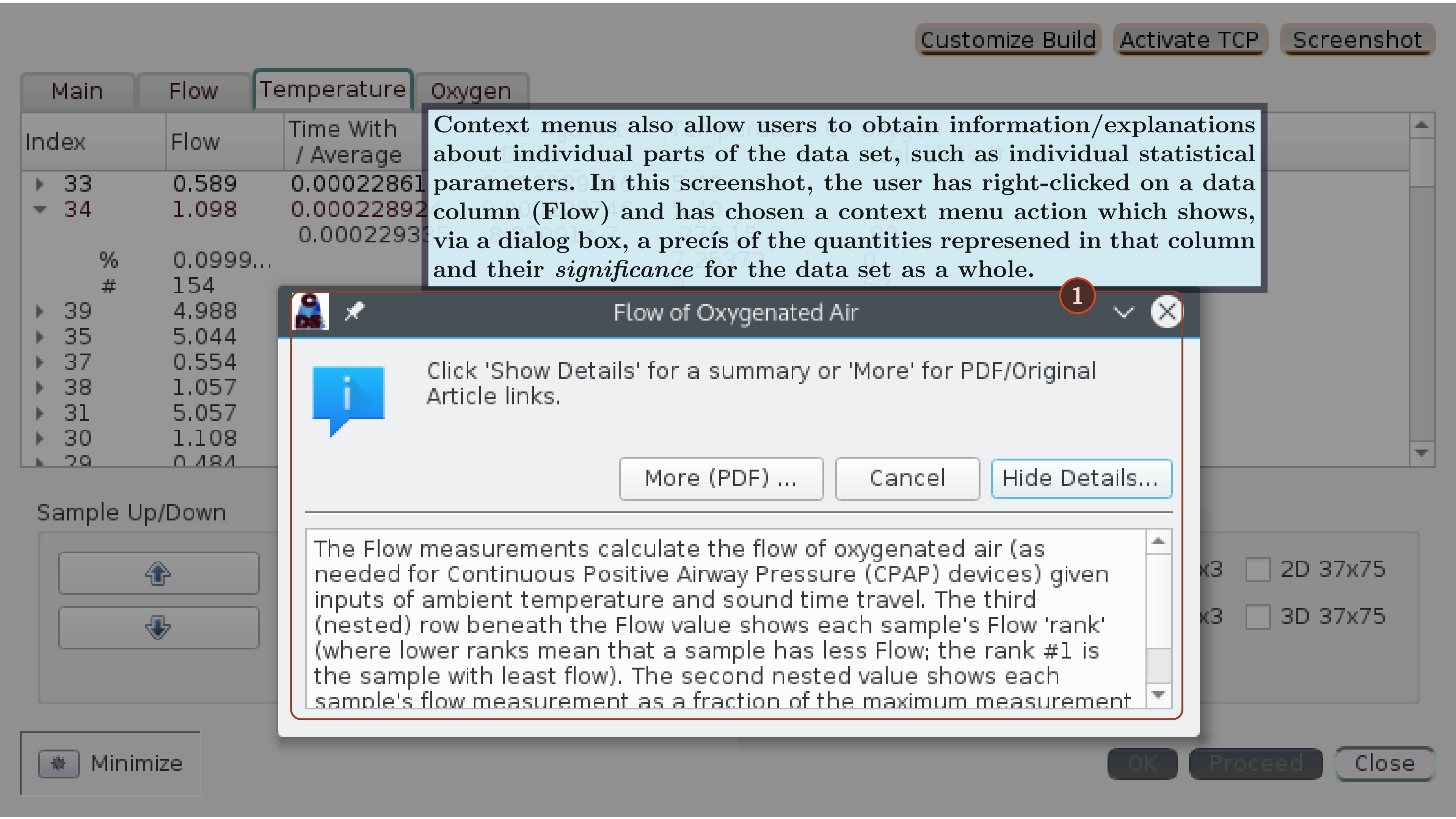
In addition to interactive visualization, Dataset Applications are useful tools for understanding experimental protocols and research methods. Within Dataset Applications, modeling units such as statistical parameters and record fields are visible in situ within a GUI — identified by labels, buttons, and other interactive micro-controls. As a result, users encounter modeling elements in a structured visual-interactive 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 terms and parameters.

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

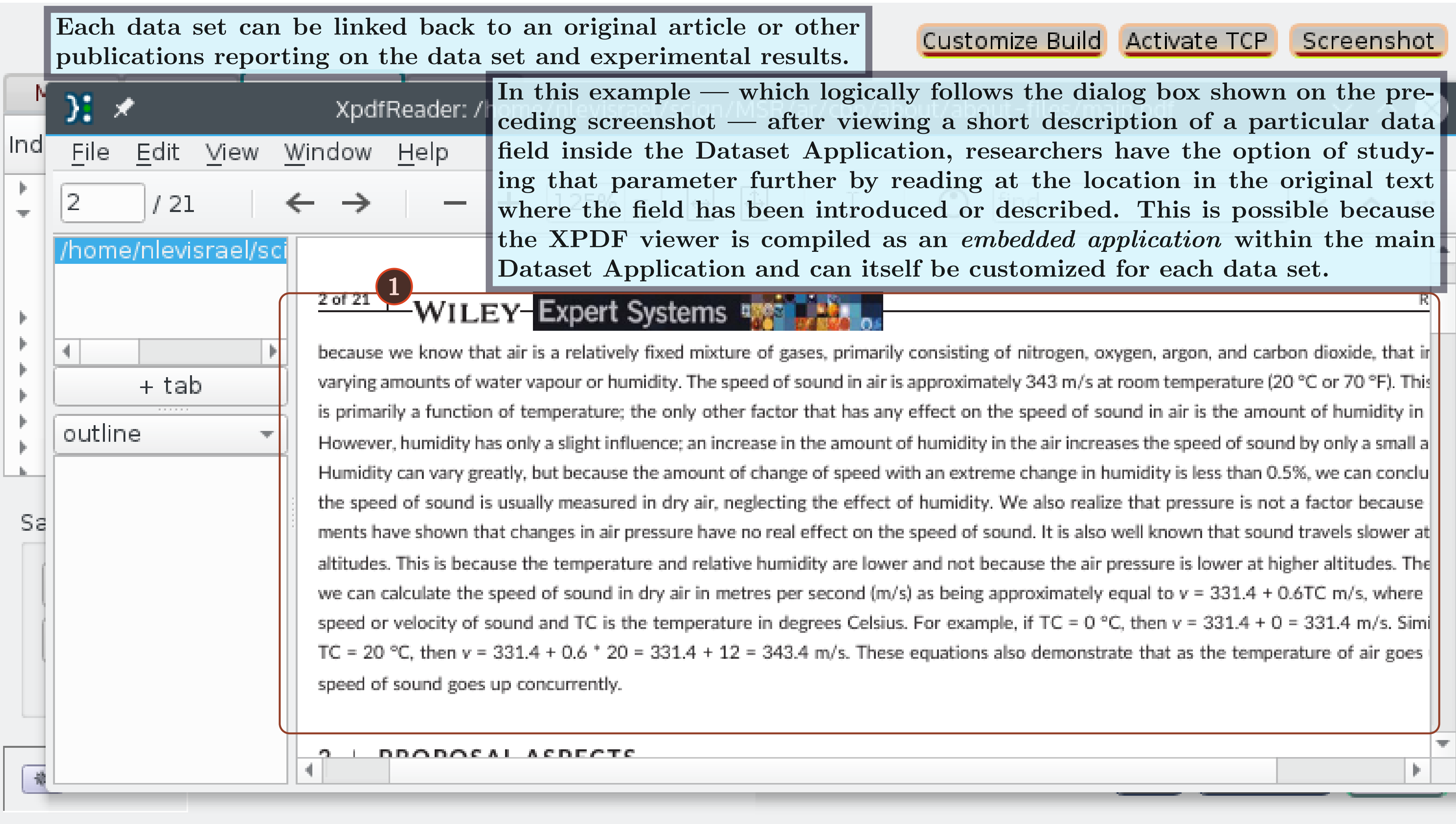
XPDF Enhancements The XPDF viewer can be customized for each data set and included with dataset code, with extra features to integrate article or book texts with Dataset Applications.

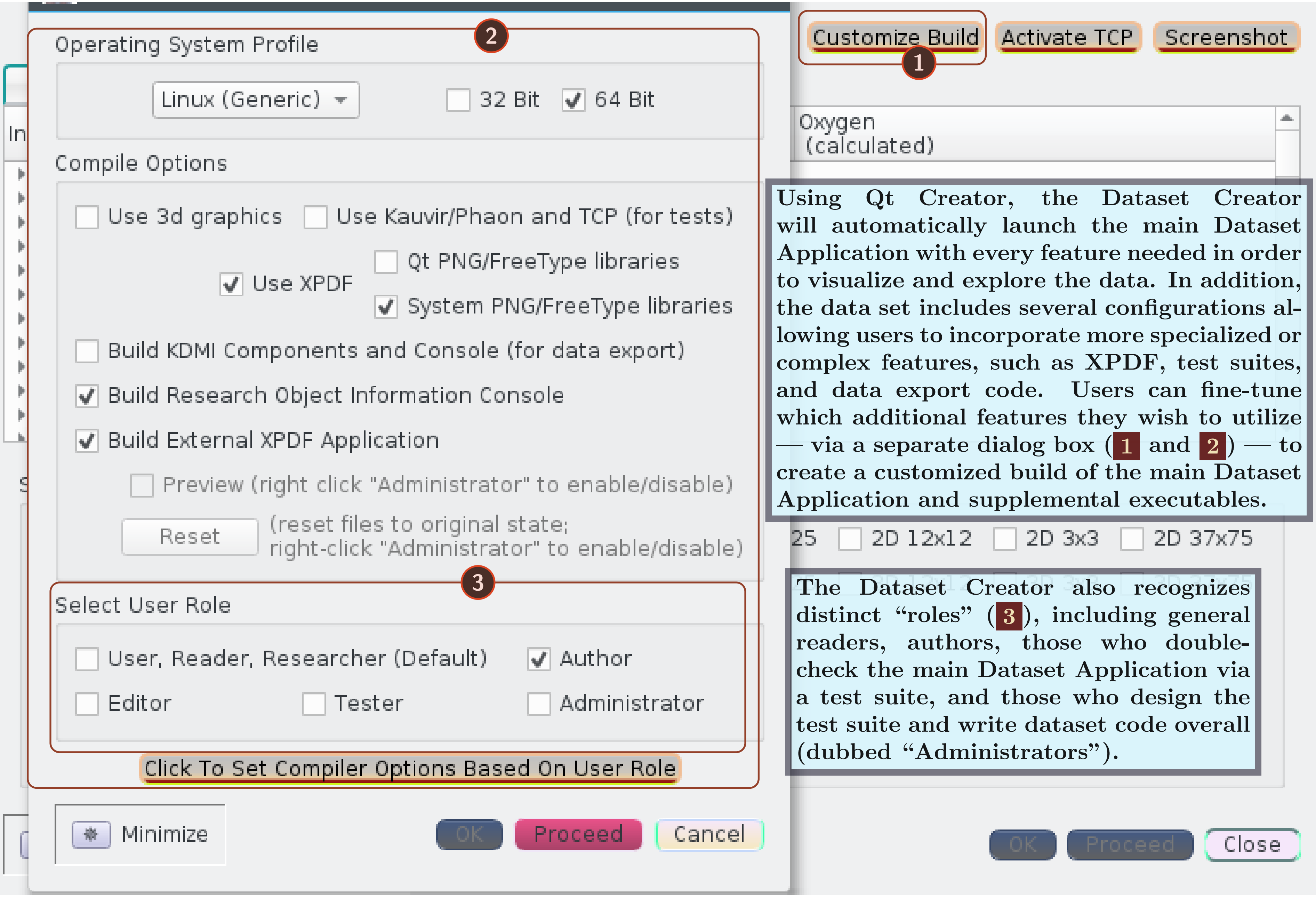
Context menus also allow users to obtain information/explanations about individual parts of the data set, such as individual statistical parameters. In this screenshot, the user has right-clicked on a data column (Flow) and has chosen a context menu action which shows, via a dialog box, a precís of the quantities represened in that column and their *significance* for the data set as a whole.



Each data set can be linked back to an original article or other publications reporting on the data set and experimental results.

In this example — which logically follows the dialog box shown on the preceding screenshot — after viewing a short description of a particular data field inside the Dataset Application, researchers have the option of studying that parameter further by reading at the location in the original text where the field has been introduced or described. This is possible because the XPDF viewer is compiled as an *embedded application* within the main Dataset Application and can itself be customized for each data set.




Operating System Profile
Linux (Generic)
32 Bit
64 Bit
Compile Options
Use 3d graphics
Use Kauvir/Phaon and TCP (for tests)
Qt PNG/FreeType libraries
Use XPDF
System PNG/FreeType libraries
Build KDMI Components and Console (for data export)
Build Research Object Information Console
Build External XPDF Application
Preview (right click "Administrator" to enable/disable)
Reset
(reset files to original state; right-click "Administrator" to enable/disable)
Select User Role
User, Reader, Researcher (Default)
Author
Editor
Tester
Administrator
Click To Set Compiler Options Based On User Role
Minimize
OK
Proceed
Cancel
Customize Build
Activate TCP
Screenshot
Oxygen (calculated)
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.
25
2D 12x12
2D 3x3
2D 37x75
The Dataset Creator also recognizes distinct “roles” (3), including general readers, authors, those who double-check the main Dataset Application via a test suite, and those who design the test suite and write dataset code overall (dubbed “Administrators”).
OK
Proceed
Close

Testing and Fine-Tuning Dataset Applications

Tools for Editors and Developers

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

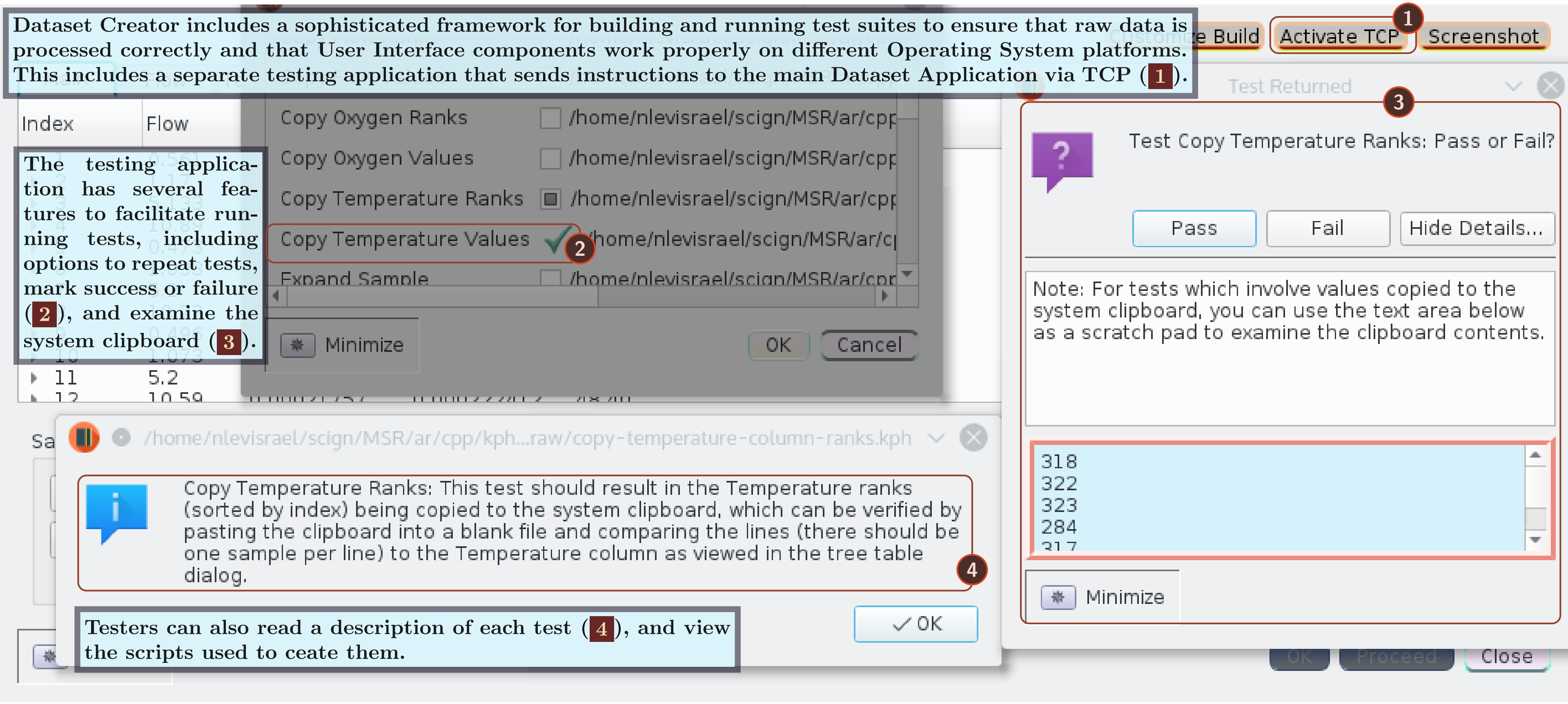
Test Suites Tools for creating and/or running test suites to ensure that the Dataset Application works across platforms.

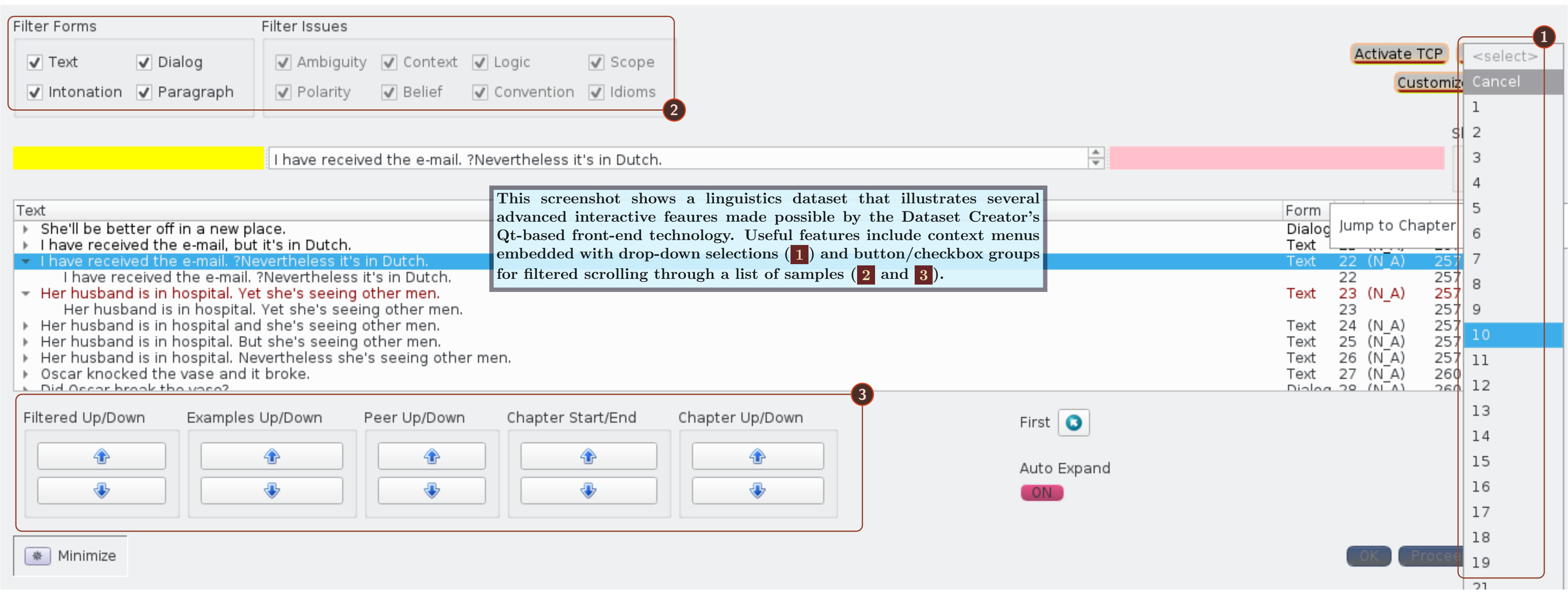
Data Export Tools for reusing data in other projects.

External Libraries Some features like XPDF and 3D graphics require libraries which are external dependencies (they cannot be published in source code form within the data set code). Advanced users can select which of these libraries to incorporate into their version of the Dataset Application.

Scripting Data sets can compile 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).


Dataset Creator includes a sophisticated framework for building and running test suites to ensure that raw data is processed correctly and that User Interface components work properly on different Operating System platforms. This includes a separate testing application that sends instructions to the main Dataset Application via TCP (1).
The testing application has several features to facilitate running tests, including options to repeat tests, mark success or failure (2), and examine the system clipboard (3).
Testers can also read a description of each test (4), and view the scripts used to ceate them.
Activate TCP
Screenshot
Index
Flow
Copy Oxygen Ranks
Copy Oxygen Values
Copy Temperature Ranks
Copy Temperature Values
Minimize
OK
Cancel
Test Returned
Test Copy Temperature Ranks: Pass or Fail?
Pass
Fail
Hide Details...
Note: For tests which involve values copied to the system clipboard, you can use the text area below as a scratch pad to examine the clipboard contents.
318
322
323
284
Minimize
/home/nlevisrael/scign/MSR/ar/cpp/kph...raw/copy-temperature-column-ranks.kph
Copy Temperature Ranks: This test should result in the Temperature ranks (sorted by index) being copied to the system clipboard, which can be verified by pasting the clipboard into a blank file and comparing the lines (there should be one sample per line) to the Temperature column as viewed in the tree table dialog.
OK
Close



This screenshot shows a linguistics dataset that illustrates several advanced interactive feaures made possible by the Dataset Creator's Qt-based front-end technology. Useful features include context menus embedded with drop-down selections (1) and button/checkbox groups for filtered scrolling through a list of samples (2 and 3).

Features of Dataset Applications for Books

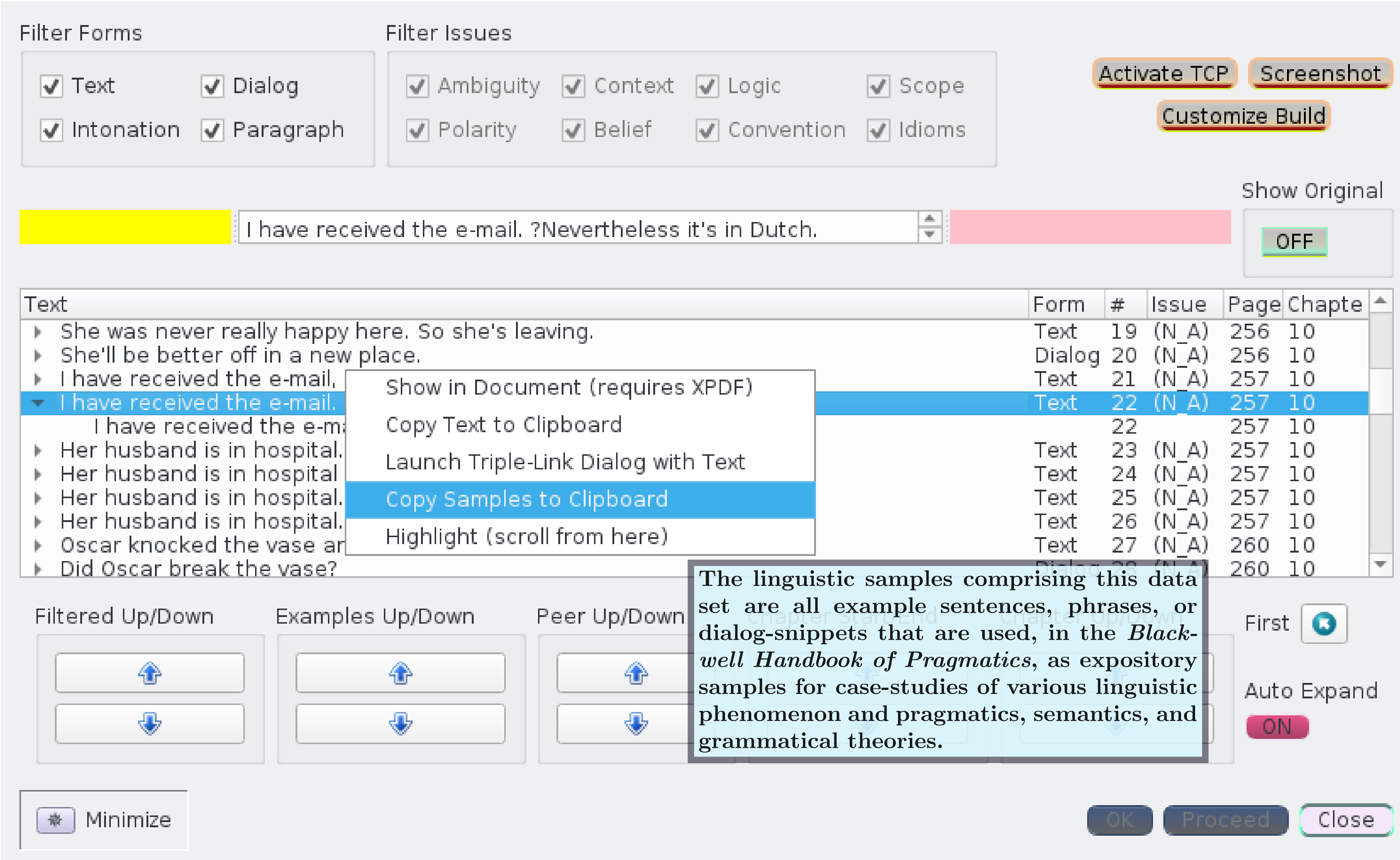
Datasets Compiled From Book Examples

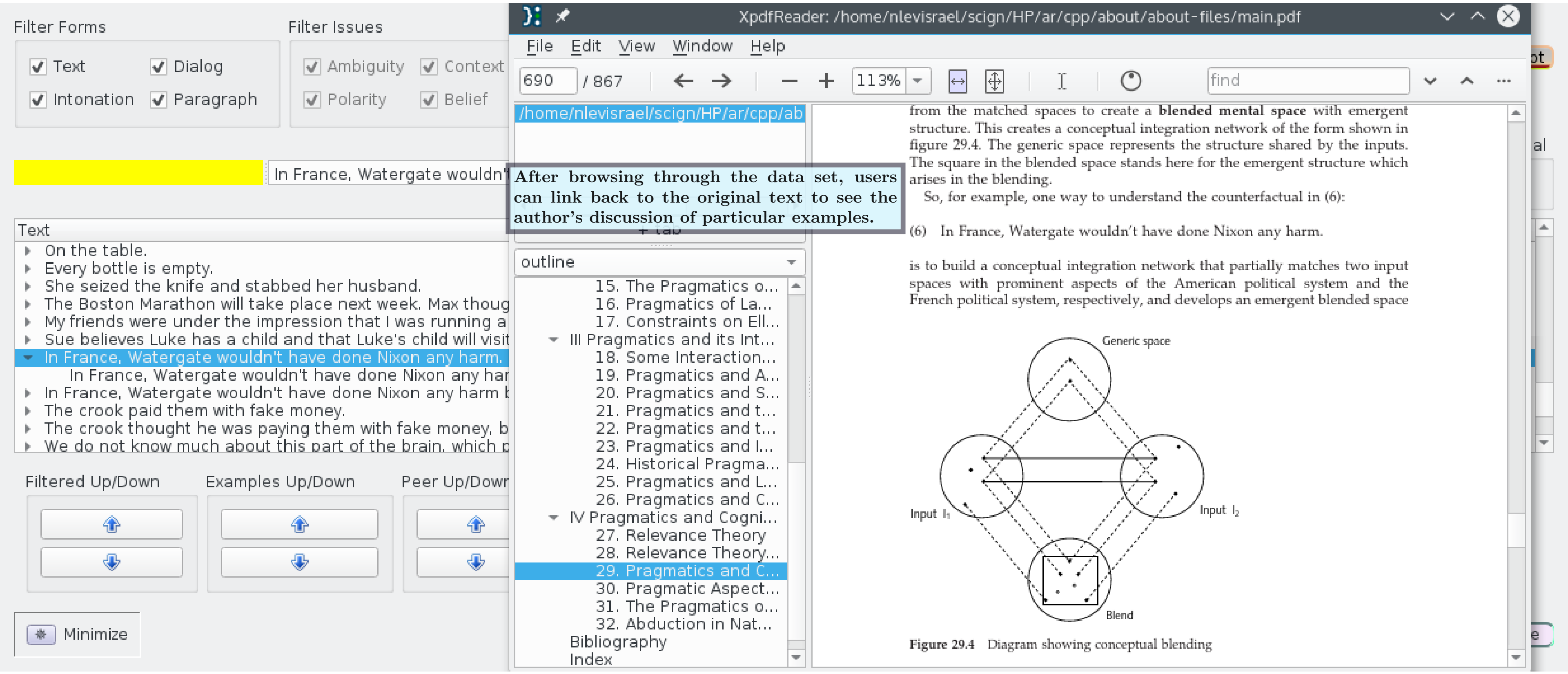
The remaining screenshots demonstrate how data sets can be used even outside the context of generating experimental data. The pictured data set represents a corpus of linguistic examples mined from Wiley's *Blackwell Handbook of Pragmatics*. Creating data sets from book-length publications can encompass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

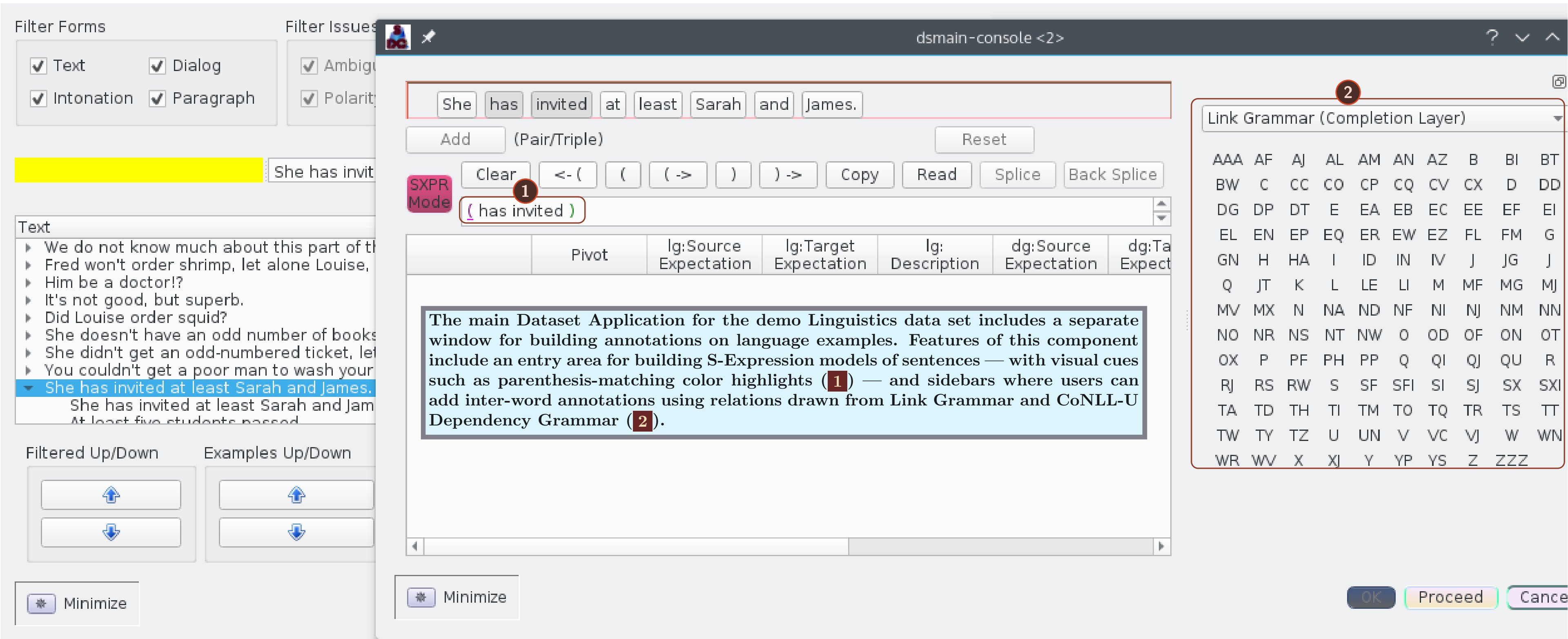
Text Mining In the case of linguistics, this involves locating example sentences within linguistics texts and storing them as an independent corpus.

Canonical Formatting If possible, linguistics texts should be formatted with markup allowing examples to be extracted automatically. This has the added benefit of ensuring that the dataset software can link between individual samples and their location in the book text.

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


Filter Forms
Text
Dialog
Intonation
Paragraph
Filter Issues
Ambiguity
Context
Logic
Scope
Polarity
Belief
Convention
Idioms
Activate TCP
Screenshot
Customize Build
Show Original
I have received the e-mail. ?Nevertheless it's in Dutch.
OFF
Text
Form
#
Issue
Page
Chapte
She was never really happy here. So she's leaving.
Text 19 (N_A) 256 10
She'll be better off in a new place.
Dialog 20 (N_A) 256 10
I have received the e-mail.
Text 21 (N_A) 257 10
I have received the e-mail.
Text 22 (N_A) 257 10
22 257 10
Her husband is in hospital.
Text 23 (N_A) 257 10
Her husband is in hospital
Text 24 (N_A) 257 10
Her husband is in hospital.
Text 25 (N_A) 257 10
Her husband is in hospital.
Text 26 (N_A) 257 10
Text 27 (N_A) 260 10
Did Oscar break the vase?
260 10
Show in Document (requires XPDF)
Copy Text to Clipboard
Launch Triple-Link Dialog with Text
Copy Samples to Clipboard
Highlight (scroll from here)
The linguistic samples comprising this data set are all example sentences, phrases, or dialog-snippets that are used, in the Blackwell Handbook of Pragmatics, as expository samples for case-studies of various linguistic phenomenon and pragmatics, semantics, and grammatical theories.
Filtered Up/Down
Examples Up/Down
Peer Up/Down
First
Auto Expand
ON
Minimize
OK
Proceed
Close


After browsing through the data set, users can link back to the original text to see the author's discussion of particular examples.
Filter Forms
Text
Dialog
Intonation
Paragraph
Filter Issues
Ambiguity
Context
Polarity
Belief
In France, Watergate wouldn't
Text
On the table.
Every bottle is empty.
She seized the knife and stabbed her husband.
The Boston Marathon will take place next week. Max thoug
My friends were under the impression that I was running a
Sue believes Luke has a child and that Luke's child will visit
In France, Watergate wouldn't have done Nixon any harm.
In France, Watergate wouldn't have done Nixon any har
In France, Watergate wouldn't have done Nixon any harm b
The crook paid them with fake money.
The crook thought he was paying them with fake money, b
We do not know much about this part of the brain, which p
Filtered Up/Down
Examples Up/Down
Peer Up/Dowr
Minimize
XpdfReader: /home/nlevisrael/scign/HP/ar/cpp/about/about-files/main.pdf
File
Edit
View
Window
Help
690
/ 867
113%
find
/home/nlevisrael/scign/HP/ar/cpp/ab
outline
15. The Pragmatics o...
16. Pragmatics of La...
17. Constraints on Ell...
III Pragmatics and its Int...
18. Some Interaction...
19. Pragmatics and A...
20. Pragmatics and S...
21. Pragmatics and t...
22. Pragmatics and t...
23. Pragmatics and I...
24. Historical Pragma...
25. Pragmatics and L...
26. Pragmatics and C...
IV Pragmatics and Cogni...
27. Relevance Theory
28. Relevance Theory...
29. Pragmatics and C...
30. Pragmatic Aspect...
31. The Pragmatics o...
32. Abduction in Nat...
Bibliography
Index
from the matched spaces to create a **blended mental space** with emergent structure. This creates a conceptual integration network of the form shown in figure 29.4. The generic space represents the structure shared by the inputs. The square in the blended space stands here for the emergent structure which arises in the blending.
So, for example, one way to understand the counterfactual in (6):
(6) In France, Watergate wouldn't have done Nixon any harm.
is to build a conceptual integration network that partially matches two input spaces with prominent aspects of the American political system and the French political system, respectively, and develops an emergent blended space
Generic space
Input I1
Input I2
Blend
Figure 29.4 Diagram showing conceptual blending


Filter Forms
Text
Dialog
Intonation
Paragraph
Filter Issues
dsmain-console <2>
She has invited at least Sarah and James.
Add
(Pair/Triple)
Reset
SXPR Mode
Clear
Copy
Read
Splice
Back Splice
(has invited)
Pivot
lg:Source Expectation
lg:Target Expectation
lg: Description
dg:Source Expectation
Link Grammar (Completion Layer)
The main Dataset Application for the demo Linguistics data set includes a separate window for building annotations on language examples. Features of this component include an entry area for building S-Expression models of sentences — with visual cues such as parenthesis-matching color highlights (1) — and sidebars where users can add inter-word annotations using relations drawn from Link Grammar and CoNLL-U Dependency Grammar (2).
We do not know much about this part of th
Fred won't order shrimp, let alone Louise,
Him be a doctor!?
It's not good, but superb.
Did Louise order squid?
Filtered Up/Down
Examples Up/Down
Minimize
OK
Proceed

A Linguistics Annotation System

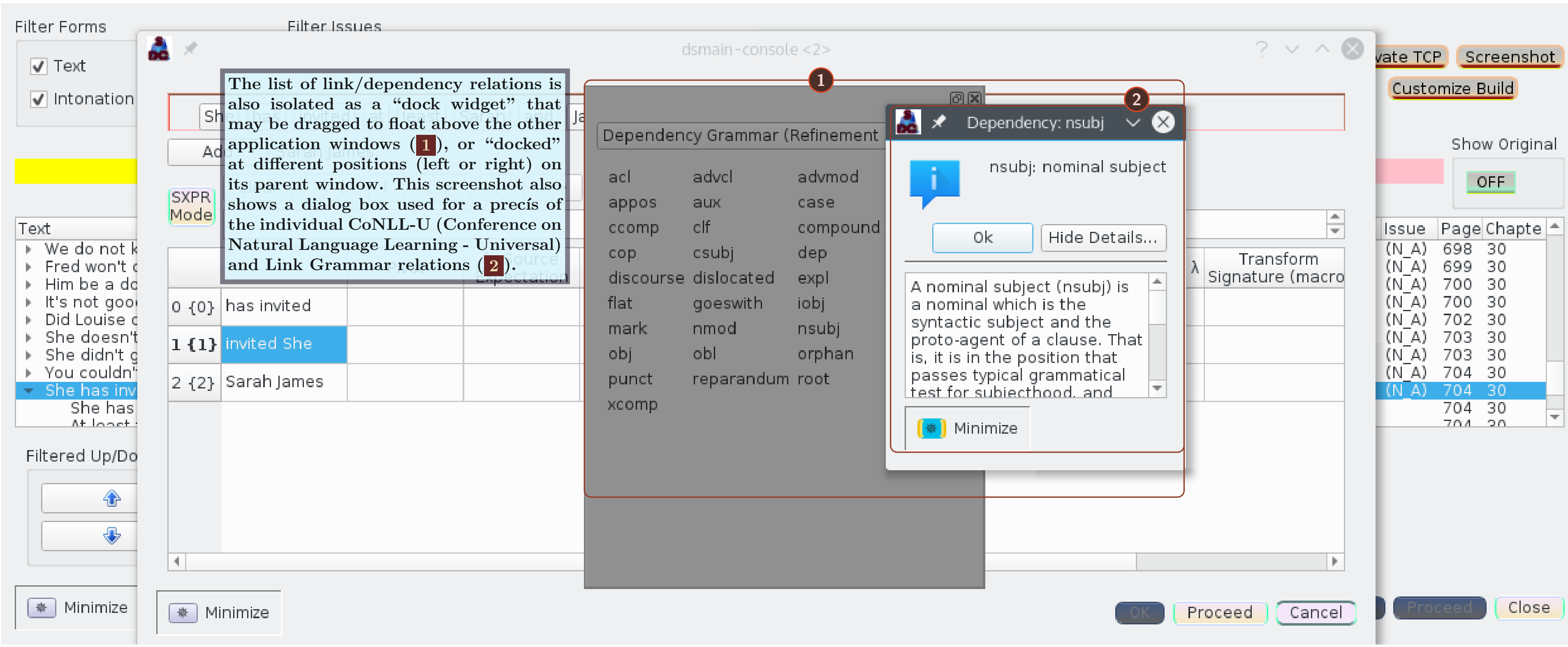
Tools to Facilitate Annotating Linguistic Corpora

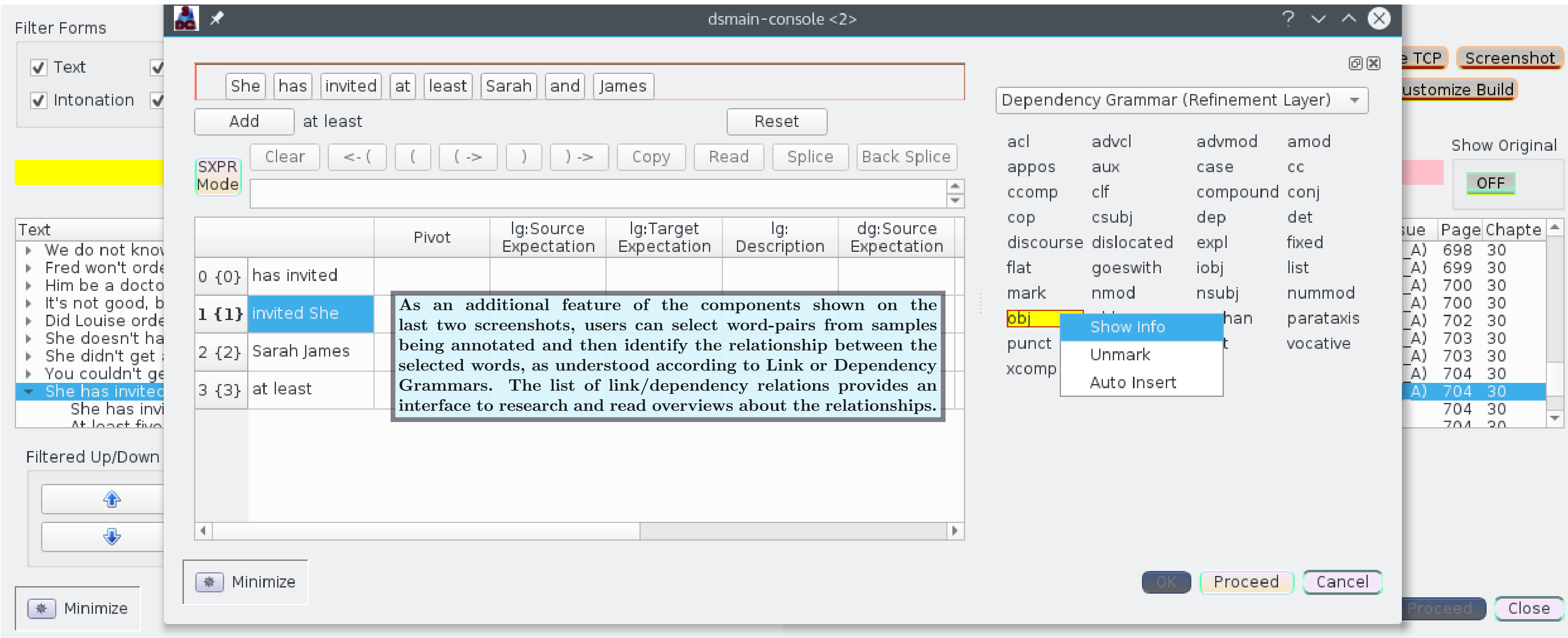
The final three screenshots show an example of how a custom-designed application can facilitate the task of building an annotated corpus from a linguistics text. The components demonstrated here enable several strategies (which can be combined) for describing parsing structures and the logical forms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

S-Expressions Representing linguistic units as semantic and syntactic transformations triggered by words assigned to "functional" (lexical or Part of Speech) types.

Dependency Grammar Representing phrase structures via inter-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, connectors represent how one word's meaning or grammatical contribution can be "completed" by the phenomenon of its linking to another word.


The list of link/dependency relations is also isolated as a "dock widget" that may be dragged to float above the other application windows (1), or "docked" at different positions (left or right) on its parent window. This screenshot also shows a dialog box used for a precís of the individual CoNLL-U (Conference on Natural Language Learning - Universal) and Link Grammar relations (2).
Filter Forms
Filter Issues
Text
Intonation
dsmain-console <2>
Dependency Grammar (Refinement
acl advcl advmod
appos aux case
ccomp clf compound
cop csubj dep
discourse dislocated expl
flat goeswith iobj
mark nmod nsubj
obj obl orphan
punct reparandum root
xcomp
Dependency: nsubj
nsubj: nominal subject
Ok
Hide Details...
A nominal subject (nsubj) is a nominal which is the syntactic subject and the proto-agent of a clause. That is, it is in the position that passes typical grammatical test for subjecthood, and
Minimize
0 {0} has invited
1 {1} invited She
2 {2} Sarah James
SXPR Mode
Transform Signature (macro
Show Original
OFF
Screenshot
Customize Build
Issue Page Chapte
Filtered Up/Do
Proceed
Cancel
Close


dsmain-console <2>
She has invited at least Sarah and James
Add
at least
Reset
SXPR Mode
Clear
<- (
(
(->
)
) ->
Copy
Read
Splice
Back Splice
Pivot
lg:Source Expectation
lg:Target Expectation
lg: Description
dg:Source Expectation
0 {0} has invited
1 {1} invited She
2 {2} Sarah James
3 {3} at least
As an additional feature of the components shown on the last two screenshots, users can select word-pairs from samples being annotated and then identify the relationship between the selected words, as understood according to Link or Dependency Grammars. The list of link/dependency relations provides an interface to research and read overviews about the relationships.
Dependency Grammar (Refinement Layer)
acl
advcl
advmod
amod
appos
aux
case
cc
ccomp
clf
compound
conj
cop
csubj
dep
det
discourse
dislocated
expl
fixed
flat
goeswith
iobj
list
mark
nmod
nsubj
nummod
obj
parataxis
punct
vocative
xcomp
Show Info
Unmark
Auto Insert
Minimize
OK
Proceed
Cancel
Filter Forms
Text
Intonation
Filtered Up/Down
Screenshot
Show Original
OFF
Page
Close