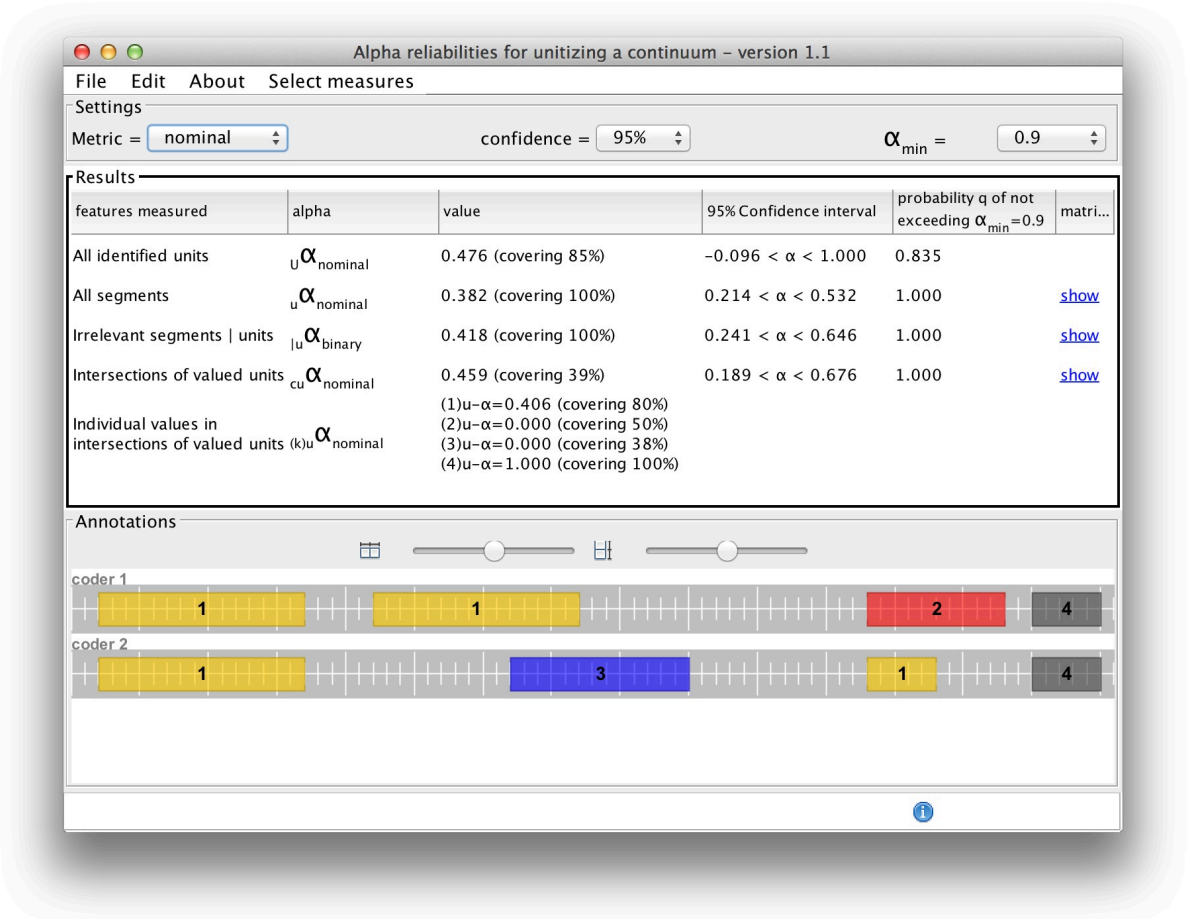
Krippendorff's Alpha-reliabilities for Unitizing a Continuum

Software Users' Manual



Date: 2016-11-29

Written by Yann Mathet yann.mathet@unicaen.fr
In consultation with Klaus Krippendorff kkrippendorff@asc.upenn.edu

0. Purpose

This u-Alpha software, computes several agreement coefficients to evaluate the reliability of segmenting, unitizing and coding continuous phenomena – texts, videos, conversations.

It is based on publications primarily by

Krippendorff, Klaus; Mathet, Yann; Bouvry, Stéphane & Widlöcher, Antoine (in press). On the Reliability of Unitizing a Continuum, Further Developments, *Quantity and Quality*.

And secondarily by

Krippendorff, Klaus (2013) *Content Analysis ; An Introduction to Its Methodology, 3rd ed., 2nd printing*. Thousand Oaks, CA: Sage Publications.

1. Requirements

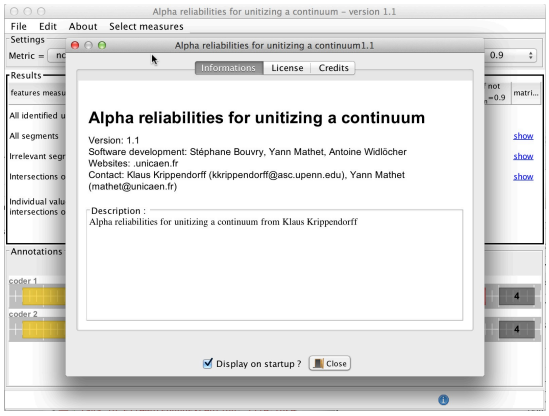
This software is written in Java (for version 6 or higher). It can be run on any system (including Mac, Windows, Linux). If Java is not yet installed on your computer or not current, you will have to install the latest version (see next section "Getting started").

2. Getting started

- Download the file named "**uAlpha.zip**" either from http://web.asc.upenn.edu/usr/krippendorff/dogs.html or from https://mathet.users.greyc.fr/agreement/
- Unzip this file (using winrar or equivalent) to get the folder "uAlphaSoft"
- Open the folder "uAlphaSoft", then the folder "software", and double-click on "uAlpha.jar"
- The software should launch and a start window should appear. IMPORTANT: recent systems have more and more restrictive security policies, and may prevent you from launching the software because it was downloaded from the web:

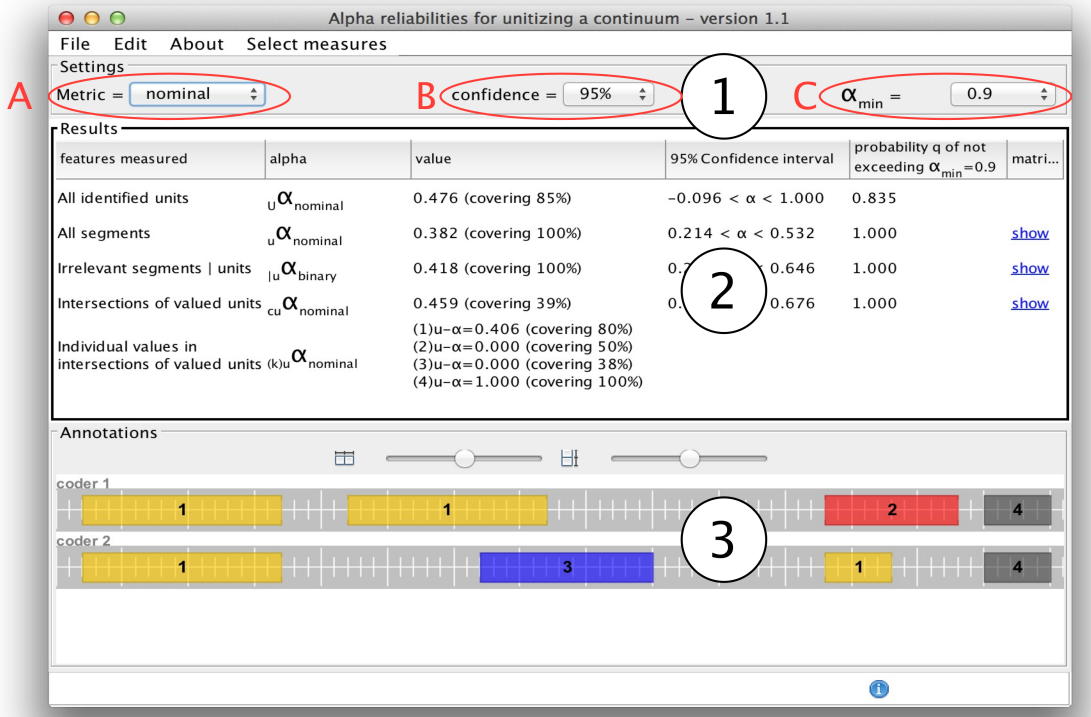
Would it happen, go to your security options (on Mac Os X: System preferences, then security and confidentiality, then in the first tab, you should get a message about uAlpha.jar, and click on the button to authorize it).

- If the software does not launch after double-clicking on "uAlpha.jar", Java needs to be installed or updated, for example from http://www.oracle.com/technetwork/java/javase/downloads/index.html. Then double-click again on "uAlpha.jar."
- Before accessing the user interface, you need to affirm the user agreement from the start window, which also contains tabs for information to consult and credits:



- You can get this window back at any moment (while using the software) by clicking on the "About" menu of the interface.

3. Introducing the interface



When you launch the software, it automatically opens with the example shown in Figure 12.8, p.311 from (Krippendorff, 2013) – cited in **section 0**.

The interface consists of four parts. The menu at the top, and three panels numbered 1 to 3 on the above screenshot:

- The menu from where you can select 3 actions: "File" to upload your data, "About" to get the credits, and "Select measures" you want to compute.

- Panel 1 is the setting panel. You can play with the settings of three variables: the applied metric in part A of the panel, the confidence level in part B of the panel, and the min-Alpha value in part C of the panel. They will be detailed in **section 4**
- Panel 2 will give you the results of the unitizing alpha coefficients that you have selected. They will be described in **section 5**
- Panel 3 depicts the loaded annotations. It is automatically zoomed to fit exactly the length of the continuum (76 in the example), with one named line per coder (two coders in the example), and colors for unlike valuations of identified units. You may adjust the horizontal and vertical zooms of this view by moving the two cursors just above the annotations.

4. Settings on panel 1

The software offers three variables:

A- Metrics: "no metric", "nominal" (by default), "interval", "ordinal" and "ratio". The chosen metric needs to reflect the level of measurement of the units' valuation. The selected metric is applied to each compliant measure.

B- Confidence levels can be chosen among: 99%, 98%, 95% (default) and 90%. The chosen level determines the 4th column of the results. For instance, in the example, u-Alpha-nominal has 95% of chance to be within 0.251 and 0.532.

C- The targeted minimum value of alpha can be chosen among: 0.6, 0.667, 0.7, 0.8, and 0.9. This choice determines the 5th column of the results. In the example, u-Alpha is well below the targeted minimum alpha of 0.9; hence the probability of not reaching this target is 1.0.

5. Results in panel 2

The results are tabulated in panel 2. You will find six columns for each measure:

1st A brief phrase explaining of what the coefficient addresses

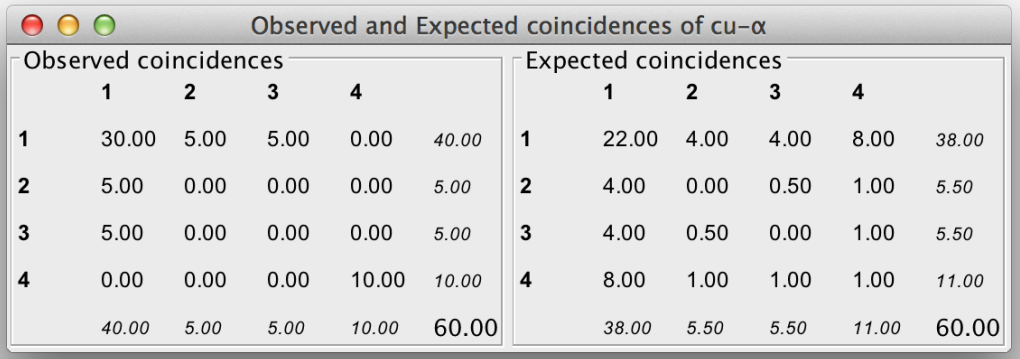
2nd Its name and the applied metric

3rd The value(s) of the coefficient and the percent of the continuum covered by it

4th Its confidence interval on the level chosen in panel 1

5th Its probability *q* of not reaching the targeted minimum alpha

6th Where available, clicking on "Show" will get you to the observed and expected coincidence matrices representing the reliability data from which the measure is computed. The screenshot below shows the matrices of cu-alpha you will see with the default example of the software. You can see for instance that the total length of intersections of units valued "1" and units valued "2" is observed to be 5 (cells 1-2 and 2-1 of the observed coincidences), and expected to be 4 (cells 1-2 and 2-1 of the expected coincidences), also called chance.



	1	2	3	4	
1	30.00	5.00	5.00	0.00	*40.00*
2	5.00	0.00	0.00	0.00	*5.00*
3	5.00	0.00	0.00	0.00	*5.00*
4	0.00	0.00	0.00	10.00	*10.00*
	40.00	*5.00*	*5.00*	*10.00*	60.00

	1	2	3	4	
1	22.00	4.00	4.00	8.00	*38.00*
2	4.00	0.00	0.50	1.00	*5.50*
3	4.00	0.50	0.00	1.00	*5.50*
4	8.00	1.00	1.00	1.00	*11.00*
	38.00	*5.50*	*5.50*	*11.00*	60.00

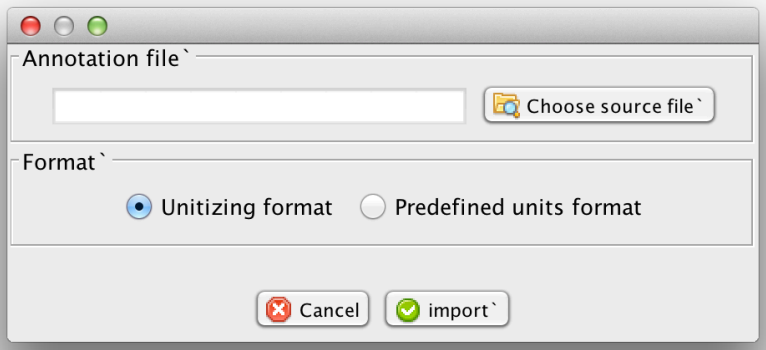
Note that confidence intervals in the 4th and the probability *q* in the 5th column are approximated by bootstrapping. Bootstrapping proceeds by randomly resampling all unitizations. Therefore, you may experience slight differences when rerunning the same data.

6. Choosing the applied measure(s) from the "Select measures" menu

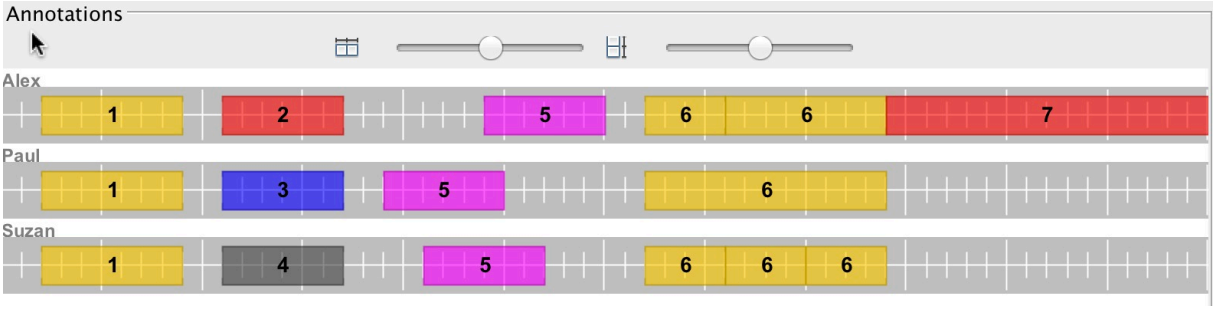
By default, the software computes and shows the results for all five unitizing alpha coefficients. However, you can opt out of computing the ones you do not wish to see. To do so, go to the "Select measures" menu on the banner, and check or un-check the measures as you please. These choices are kept in subsequent computations unless you change them.

7. Importing data from "File" menu.

- Go to "File" menu, and click on "Import annotations"
- Click on "Choose source file", and use the browser to select a file
- Select a file format: "Unitizing format" or "Predefined units format", as explained in section 8, in the center panel:



- Then click on "import"
- You can verify this by using the file provided in the distribution, in the "corpus" folder, named "**AlexPaulSuzan.csv**" – see below, using the "**unitizing**" format. You will get the example used in (Krippendorff, Mathet, Bouvry & Widlöcher, in press) see reference in **section 0**.



8. Formats of imported files

Why two formats?

This software was designed to compute the reliability of data resulting from **segmenting, unitizing and valuing units on a continuum** (where coders choose these units' length, location, and value), not from **coding predefined units** (where coders merely categorize or value given units).

However, it it is also possible to compute the c-Alpha coefficient for data resulting from coding **predefined units.** This software treats such data as if they came from a continuum in which all units are of size 1. Applying the chosen continuum-based alpha coefficients to them yields the following equivalences: When no data are missing (all units are coded by all observers), u-Alpha = cu-Alpha = c-Alpha for all appropriately chosen metrics, see (Krippendorff, 2013: 277-301) and the end of section 8.2. When data are missing, u-Alpha treats them as dummy categories. But because cu-Alpha ignores missing data relative to all intersections while c-Alpha ignores them relative to the units not coded, cu-Alpha will be more conservative than c-Alpha.

Therefore, this software recognizes two formats (to select among them see **section 7**):

- The "Unitizing format", for which this software and its reliability coefficients are designed.
- The "Predefined units format" which provides a bridge across what these two coefficients have in common.

8.1. The "Unitizing format"

IMPORTANT: the u-Alpha family was not designed to cope with overlapping units within one annotator. Hence, the software will prevent you from loading and computing such data:

All data should be comma delineated *.CSV-files, containing 4 columns, one line per unit. Their 1st column contains the name of the coder, the 2nd the value assigned to that unit, the 3rd column contains the beginning and the 4th the end of that unit. For instance, Alex's 1st unit in the preceding screenshot would have to be:

```
Alex,1,2,9
```

Herein:

`Alex` is the name of the coder of this unit.
`1` is the category or value assigned to this unit (here one numeral between 1 – 6)
`2` is the beginning of this unit, measured from the beginning of the continuum
`9` is the end of this unit, also measured from the beginning of the continuum

The complete file with its 16 units is as follows:

```
Alex,1,2,9
Alex,2,11,17
Alex,5,24,30
Alex,6,32,36
Alex,6,36,44
Alex,7,44,60
Paul,1,2,9
Paul,3,11,17
Paul,5,19,25
Paul,6,32,44
Suzan,1,2,9
Suzan,4,11,17
Suzan,5,21,27
Suzan,6,32,36
Suzan,6,36,40
Suzan,6,40,44
```

By default, the continuum:

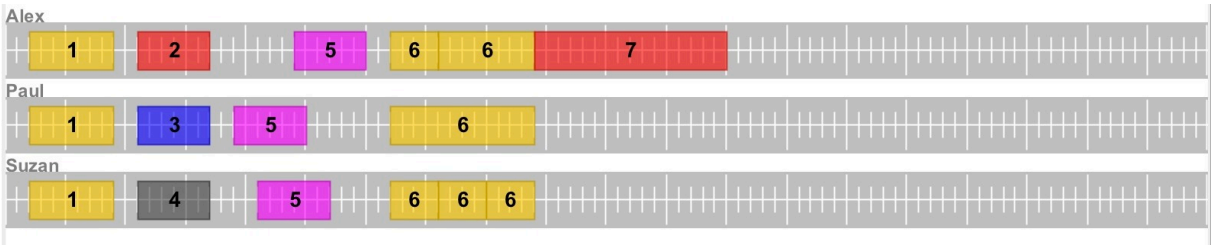
- Begins at position 0
- Ends at the highest end of anyone's last units (60 in our example, since Alex coded a unit from position 44 to position 60)

It is possible to override the default length for the continuum by adding a line preceding the list of units that specifies its length. For example:

```
length=100,0,0,0
```

This will extend the length of the continuum to force the right boundary to be 100 (instead of the default length of 60).

You can check this example with the provided file "AlexPaulSuzanLength100.csv" of the "corpus" folder, and now get the data shown below where you can see an empty part on the continuum of size 40 (i.e. 100-60):

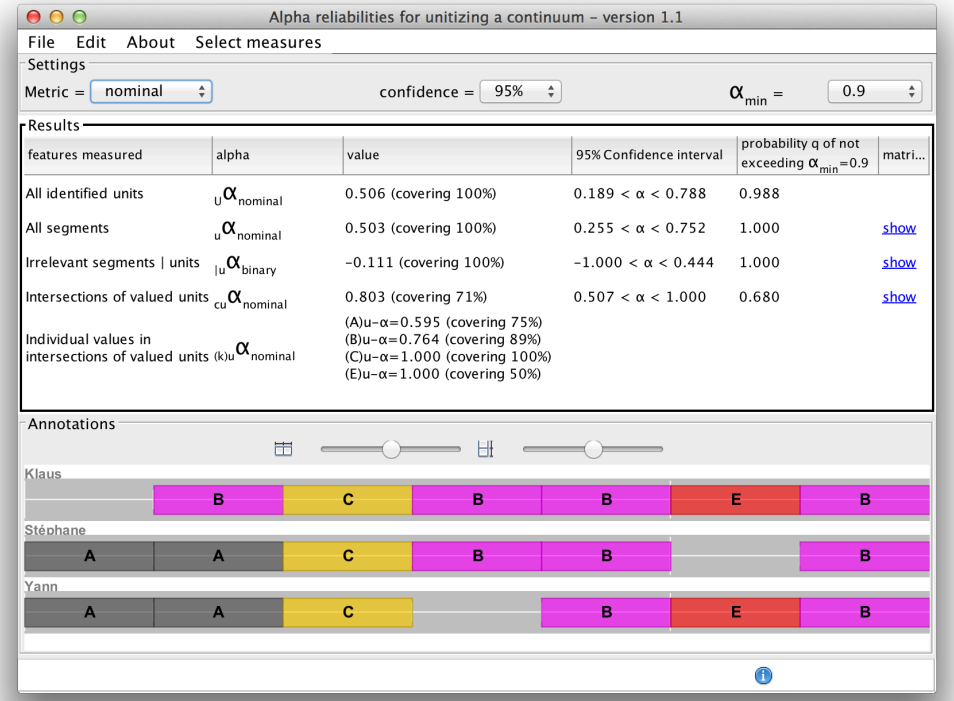


8.2. The "Predefined units format"

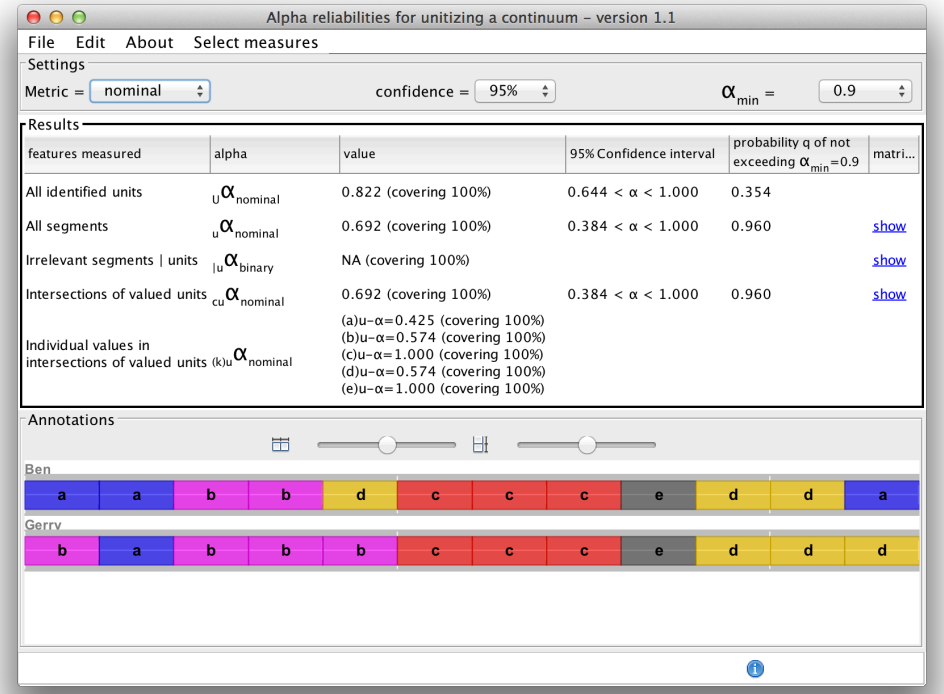
For coding predefined units, the software uses another *.CSV format where the 1st row consists of the names of the coders, followed by as many rows as there are predefined units containing the categories or values assigned by the coders. For example, the file "TestPredefinedUnits.csv" consists of seven rows:

```
Klaus,Stéphane,Yann
,A,A
B,A,A
C,C,C
B,B,
E,,E
B,B,B
```

Note various omissions: Klaus did not code the 1st unit, Yann did not code the 4th unit, and Stéphane did not code the 5th unit.



Following is an example without missing data, i.e. where all coders have coded all units, provided as "TestPredefinedUnits2011BenGerry.csv" in "corpus" folder of the distribution.



As already stated, without missing data, the u-Alpha and cu-alpha coefficients equal the c-alpha coefficient (Krippendorff, 2013: 277-301), see "Computing Krippendorff 's Alpha-Reliability", http://repository.upenn.edu/asc_papers/43 (Accessed 2015.8.13). The software moreover produces their confidence intervals and the probabilities *q* of failing to exceed a desired minimum alpha.