

An Visualized Analysis on Affix Borrowing in Global Languages

Yumo Yao
yy2650@nyu.edu
NYU Shanghai

Muqing Yang
ym2127@nyu.edu
NYU Shanghai

Siyuan Yu
sy2470@nyu.edu
NYU Shanghai

Overview

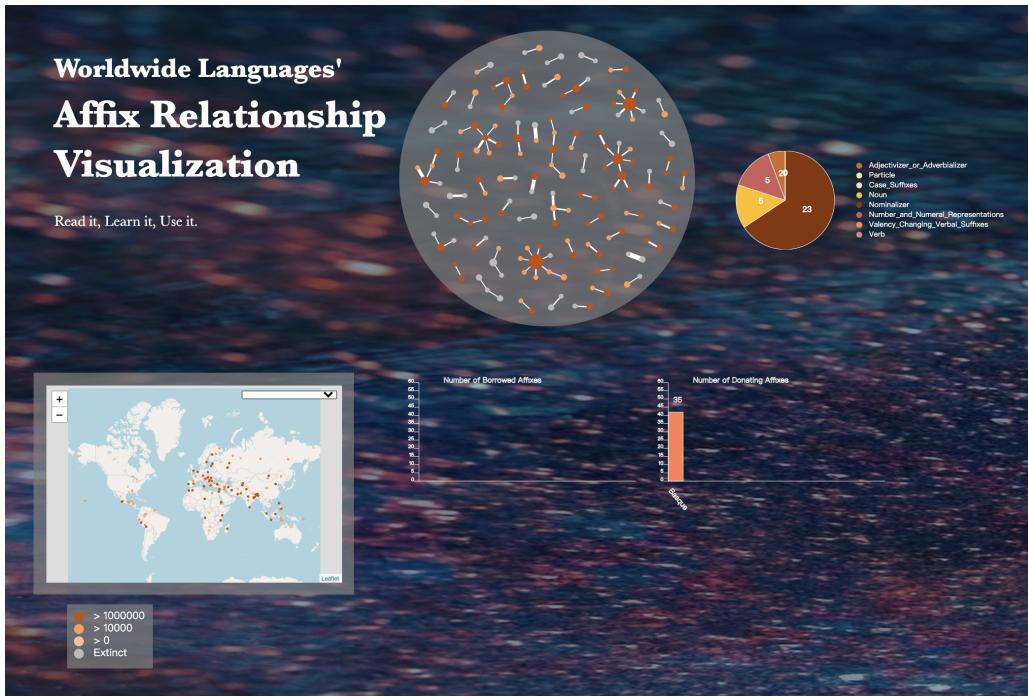


Figure 1: FG1

Affix borrowing is a normal phenomenon between languages. For example, the English affix “-able” is originally borrowed from French, meaning “being able to do” in both languages. The interactions between languages represented by affix borrowing reveal the intrinsic connections of the culture and history. Our visualization system analyzes the affix borrowing phenomenon. It shows the distribution of languages and the specific information including number of affix borrowed, affix categories in each affix-borrowing relationship.

I. Description of Data Set and Processing

We intend to use the AFBO dataset for the source of visualization. This dataset indicates the recipient-donor relationship between 2 languages as well as the details of borrowed affixes. We may also individually collect other data by researching. Our final dataset includes 2 parts:

1. language information, ie.,
 - (a) longitude, quantitative, for language's geographic location.
 - (b) latitude, quantitative, for language's geographic location.
 - (c) macro areas, categorical, for categorizing languages by geographic/cultural areas.
 - (d) number of native speakers, quantitative, for representing the language's popularity in the map and the link-node diagram.
2. Recipient-donor relationship information, i.e.,
 - (a) borrower, categorical, language names as borrower.

- (b) donor, categorical, language names as donor.
- (c) number of each kind of affixes borrowed (adjectivizer/adverbializer, particle, case suffixes, noun, nominalizer, number numeral representation, valency-changing verbal suffixes, verb), quantitative, for numbers of affixes of each category borrowed in a specific borrowing relationship.

AFBO data set URL: <https://afbo.info/download>

Data preprocessing pipeline: merge language information with recipient-donor relationship information for the node-link diagram usage. Columns include [Language_index, Language, (Language) Population, (Language) Count, Donor_index, Donor, Affix_num]

II. Goals and Tasks

Languages are essential communication tools of our life. This visualization analysis aims at visualizing the information about languages and their inner connections by affix borrowing. Our goal is to provide an intuitive visualization that is understandable to the laymen about how the languages are distributed and connected to each other. Such a goal was implemented by a map and a node-link diagram, which are straightforward in displaying the geographical position and the relationship. Also, we designed a bar chart and a pie chart to provide specific information of each affix-borrowing relationship. For example, if a user is interested in one specific language, he can zoom in the map to see where the language is spoken and its population of native speakers. He may also have a look at the node-link diagram to see how its affixes are borrowed from/ donated to any other languages. He may also have detailed information about the affixes to see the number and specific categories.

III. Description of Visualization & Sketches

Fg1 Symbol Map

In this diagram, we apply the Leaflet package to visualize the location and number of users of each language in the data set. We visualize the geometry information of each language by map metaphor (locate each language on the world map). The shape of the objects is point, and the map is used as reference. We also visualize the number of users of each language by color saturation of points. The more users a language has, the more saturated the color of the point representing this language is.

Graphical information is the most important information to be represented in the map diagram, therefore we use the position channel to encode it. Population is not so important so we use color saturation to encode it. The variation of population is large, therefore we divide population into 4 scales: Extinct, <10000 , <1000000 , $>=1000000$

Selection box: When selecting a continent, the map will zoom into the area of this continent and show the points.

Tooltip: When mouse is on a node, tooltip will show the name and population of this language.

Interaction: When a mouse clicks a node, both this

node and the corresponding node will be highlighted for several seconds.

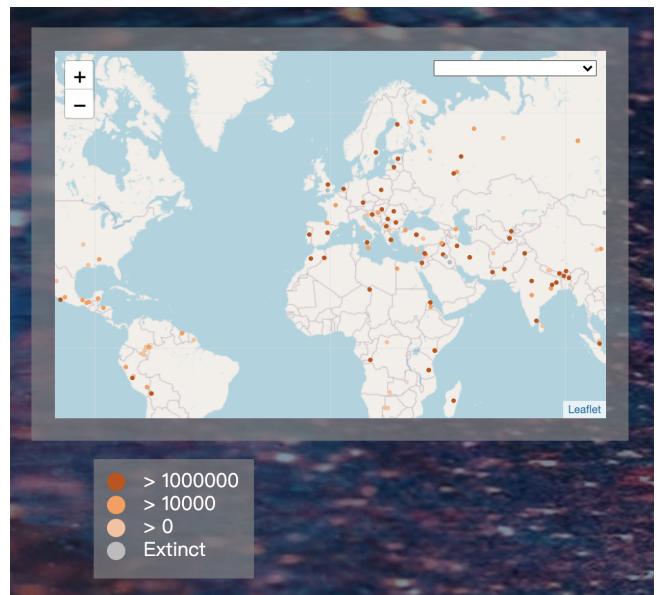


Figure 2: FG2

Fg2 Node-Link Diagram

In this diagram, we visualize the borrowing/lending relationship between affixes and the number of users in languages. The shape of the objects is point, and we visualize the number of users of each language by color saturation of points. We also visualize the times of each language being a donor of other languages through the size of points. The more recipient languages this language has, the larger the point representing the language will be. We also visualize the borrowing/lending relationship between affixes in languages by links. For each pair of borrowing/lending relationship, the two languages in this relationship are connected by links. The thickness of the link is represented the number of borrowed affixes between the two languages.

Affix relationship is the most important information to be represented in node-link diagram, therefore we use size channel to encode it (both in nodes and in links). Population is not so important so we use color saturation to encode it. The variation of population is large, therefore we divide population into 4 scales: Extinct, <10000, <1000000, >=1000000

Tooltip: When mouse is on a node, tooltip will show the name and population of this language. When mouse is on a link, tooltip will show the name of the recipient and donor languages, as well as the number of affixes in this relationship.

Selection box: When selecting a continent (in the selection box in Fg 1.), only the languages with geographical locations in the continent will be selected and displayed.

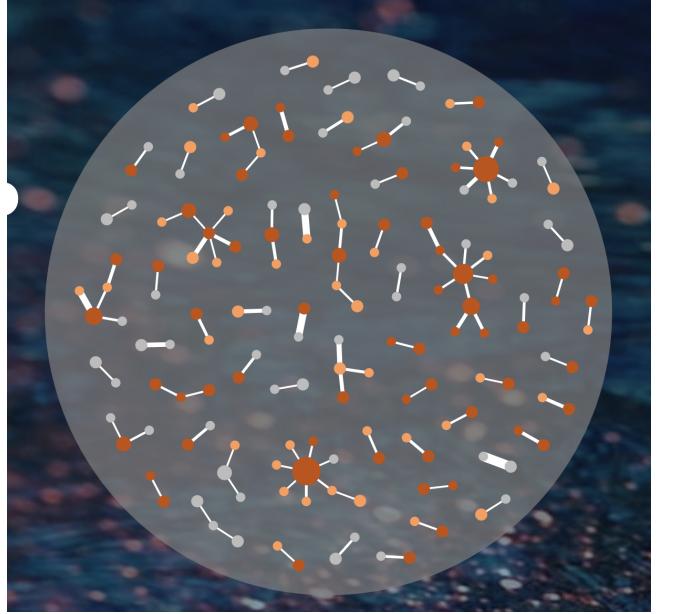


Figure 3: FG3

Fg3 Bar Charts

In this diagram, we visualize the number of lent/borrowed affixes in each language. If a point is selected (in either the map in Fg.1 or the node-link connection in Fg.2), the information regarding the affixes lent and borrowed by the language will be displayed in a form of the barchart. One bar chart depicts every borrowing relationship of the selected language: what languages it borrows affixes from and the number of affixes borrowed. Another bar chart depicts the lending relationship in the same way. Also, if a link is selected in Fg.2, the lending relationship of the donor of this language will be shown in the bar chart. The x-axis

is the language name, and the y-axis is the number of affixes. We have not used additional channels to visualize.

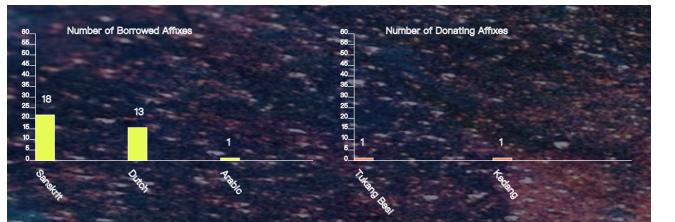


Figure 4: FG4

Fg4 Pie Chart

In this diagram, we visualize the details of borrowing/lending affixes between two languages. If a connection (edge) is selected in Fg.2, a pie chart will be shown to specifically elaborate the selected borrowing/lending relationship between two languages. That is, zooming into the connection by the type of the affixes including number, nominalizer, adjectivizer, etc. We use color channels to encode the type of affix.



Figure 5: FG5

IV. Reflection

Our project provides a clear visualization of global languages and the affix-borrowing phenomenon. Through the visualization and analysis, we found out the possible relationship between the population and the borrowed/lent affixes, as well as the relationship between the affix-borrowing relationship and the geographical location.

In our project, we generally followed our proposal, providing a comprehensive visualization system to display the information. However, there are several variance from the original proposal we have changed some content compared with the proposal.

1. In the proposal, we planned to highlight the macro area while selecting through the drop-down list. However, hiding other languages in other areas in gray will miss the information, because some of the languages in other macro areas are related to the ones in the highlighted areas. Thus, we decided to zoom in the selected area so that the users can focus on it, but he may also move the map to see languages in other areas.
2. We used the Leaflet package to draw the map. Compared with React.js or D3.js, Leaflet map has more convenient functions for zooming in and out, so that the users can find the detailed location information about where the language is spoken specifically and view the density of languages existing in some certain areas.
3. We change the shape of the link from arrow to line. This is because of the design thinking. The links are mainly short, and if we add the arrow, they will become confusing and not satisfying. Also, we have set the tooltip for the link, so it is not necessary to show the link with an arrow.
4. The size channel of the node in node-link diagram now represents the number of recipients owned by the language, instead of the number of affixes donated by the language. That's because what we want to show in the node-link diagram is the macro relationship between various languages, not a single borrowing/lending relationship's information (which is represented in the bar and pie chart). For example, there may be languages that contribute a large number of affixes to a single language, and also languages that contribute a small number of affixes to many languages. If the size of the former node is larger than that of the latter node, it does not match what the node-link diagram indicates.
5. When we click on a link in the node-link diagram, the lending relationship of the donor of this language will also be shown in the bar chart. This makes users get more information about this relationship.
6. We change the color of the points to saturated orange. That's because the world map's color generated by Leaflet is generally blue and green. We still want to make a color-blind safe visualization.
7. When clicking on a point, the graph won't highlight the points that have relationships with this point, since we found that there's no need. The node-link diagram itself is already clear enough to show the relationships.

Our visualization goals keep the same. The map and node-link were decided at the beginning, which we now still consider as strong visualization tools in showing the language information and the affix-borrowing relationships. Also, such a design can cover the intended displaying of the information.

For technical goals, as it is mentioned above, we used leaflet.js to construct the world map. We have compared leaflet.js with React to see which to use, while leaflet.js was eventually chosen due to its detailed presentation of the world map.

References

- [1] A world-wide survey of affix borrowing. [Online]. Available: <https://afbo.info/download>