



What matters in free listing? A probabilistic interpretation of the Saliency Index

Leonardo da Silva Chaves^{1,2} , André Luiz Borba do Nascimento²  and Ulysses Paulino Albuquerque^{2*} 

Received: September 26, 2018

Accepted: January 30, 2019

ABSTRACT

Free listing is a data collection technique used in different subject areas to characterize a given cultural domain. Analysis of a set of lists from a human population allows inferences to be made about the cultural saliency of the items in that domain. However, the challenge that the saliency index presents is establishing a threshold value for determining whether an item can be considered salient or not. The present analysis reveals how to determine which items of a list have non-randomly determined citation frequency and order. Monte Carlo techniques were used to create a hypothetical null scenario. The present analysis not only objectively identifies which items stand out in relation to the others, it also reveals which items can be considered idiosyncratic and how order and frequency independently influence the saliency index. The present analysis represents a useful tool for analyzing data collected through free listing. It also can contribute to understanding processes related to the cultural relevance of items and to the test future hypotheses in different areas of knowledge.

Keywords: cognitive anthropology, ethnobiology, ethnobotany, free-list, human ecology

Introduction

The free list is one of the most commonly used data collection techniques in different areas of knowledge. It is an extremely useful tool that allows a rapid survey of people's knowledge about a given cultural domain (Brewer 1985).

Although free listing presents some limitations (e.g. Quinlan 2005; Sousa *et al.* 2016; Zambrana *et al.* 2018), it has been widely used in ethnobotanical studies with different approaches. Some examples of its application can be found in the investigation of plants with pharmacological

potential (Cartaxo *et al.* 2010), the comparison of knowledge about plants in different communities (Ladio *et al.* 2007) and in the study of the structure and resilience of medical systems (Santoro *et al.* 2015).

In addition to identifying the items belonging to a given domain, the analysis of different lists in the same human population allows inferences to be made about their cultural saliency. In an etic perspective, it represents a measure of the cultural importance of the mentioned items, expressed by the relation between the frequency of citations and the order of citation of each element (Quinlan 2005).

¹ Programa de Pós-Graduação em Etnobiologia e Conservação da Natureza, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil

² Laboratório de Ecologia e Evolução de Sistemas Socioecológicos, Centro de Biociências, Departamento de Botânica, Universidade Federal de Pernambuco, 50670-901, Recife, PE, Brazil

* Corresponding author: upa677@hotmail.com



During a few decades, the frequency of each item (Borgatti 1990; Weller & Romney 1988) and the order of citation (Romney & D'Andrade 1964) were used separately to express saliency. However, Smith (1993) proposed the joint use of these two variables in the Saliency Index. The Saliency Index has since been adjusted by Smith & Borgatti (1997) and Smith *et al.* (1995), while other versions were proposed by Sutrop (2001) and Robbins *et al.* (2017), although maintaining the basic idea of combining the average position and frequency of citation of each item in determining the relative importance of items.

Calculating the Saliency Index is relatively simple, and can be done using software for analyzing free lists (e.g. Borgatti 1996; Borgatti *et al.* 2002; Pennec *et al.* 2012; Purzycki & Jamieson-Lane 2017). A saliency value is calculated for each item registered in a set of free lists, which are then ranked according to their relevance. The saliency index is very important in ethnobotanical studies because of its usefulness in identifying plants with high general use value (Lozano *et al.* 2014) or species of better quality used for specific purposes (Nunes *et al.* 2016). It is also important for prospecting for species with pharmacological potential (Leitão *et al.* 2013). The interpretation of these values, however, is quite subjective. In the analysis of a set of free lists it is difficult to establish a saliency threshold (Quinlan 2005). Weller & Romney (1988), for example, suggested that items with a citation frequency of about 75 % should be considered the most important. Quinlan (2005), on the other hand, points out that breaks are often observed in descending tabulations of Saliency Indices of items of free lists, which can be used to distinguish the most important items. According to this author, items listed prior to the first break should be considered “highly salient”, while other breaks can be used to distinguish other groups of less salient items.

Borgatti & Halgin (2013) also discussed the difficulty with interpreting values of the Saliency Index in research on “cultural domains” through free lists. These authors emphasize the need to define the boundaries of a “cultural domain”, and identify which items in a set of lists can be considered “idiosyncrasies”. In addition, these authors suggest that when analyzing descending tabulations of saliency, graphical analysis of a scree plot of frequency values can help identify breaks (or elbows) and suggest the threshold between items belonging to the cultural domain and those of little or no relevance for study (Borgatti & Halgin 2013).

Since this method is very subjective, the identification of such breaks is often quite complicated (Thompson & John 2006). To the best of our knowledge, there is no method available in the literature for the analysis of free lists that proposes an objective technique for the identification of salient items. To reduce subjectivity in interpreting the Saliency Index, we present a mathematical analysis that allows understanding which elements of a set of free lists

have a frequency and order of citation that are not due to random reasons. To achieve this we use the probability of the occurrence of the saliency value of each item in a hypothetical null scenario.

Materials and methods

The “p-value” of the Saliency Index

When we claim that certain elements mentioned in the free lists are more salient than others are, we are assuming that some of these elements are cited at greater frequencies and/or with a different order than the expected by random reasons. Thus, the saliency values calculated for these elements, in addition to being higher, are not expected to occur associated with randomly cited items.

Assuming this, in order to decide on the statistical significance of the saliency value of the items of a free list, it is necessary to initially set a null scenario that presents the same characteristics of the real scenario studied. In order to achieve that we used the number of informants interviewed in the population, the total number of items cited and the average size of the free lists, to randomly generate free lists of 1000 simulated populations, using the Monte Carlo Techniques (Robert & Casella 2010). Each simulated population has the same number of items and informants as the actual one, although the frequency and order of the items in the lists are completely random. For each simulated population, we performed the calculation of the Saliency Index of each item, creating a null distribution with the saliency values of the items cited due to random reasons.

Then, from the data collected from the real population, we calculated the Saliency Index of each item cited followed by the probability of occurrence of these values in a null scenario (*p-value*). We accepted as significant all items that had lower values than the 5% probability threshold set (*p-value* < 0.05). All analysis was performed using R development environment (R Core Team 2017) and the script developed is available as supplementary material online (R Script S1 in supplementary material), or available at request.

Results

The cut-off point of Saliency Index: Using the new method

In order to exemplify the suggested analysis, we used free lists referring to the knowledge of medicinal plants in “Horizonte” community, located in the surroundings of the National Forest of Araripe-Apodi (Ceará – Northeast of Brazil). These lists are part of two already published papers (Lozano *et al.* 2014; Nascimento *et al.* 2016).



The botanical identification of the plants listed through common names on Table 1 is available in Lozano *et al.* (2014).

The free-list technique was used to investigate the knowledge about medicinal plants of the communities, having as the generative question: “Which medicinal plants do you know?” The shortest list in our subsample has three items, the largest of which is 63, and a total of 216 items have been cited along these 153 lists.

For each simulated population, we created 153 other lists that followed the parameters observed in the original lists. The size of each of the 153 lists was randomly set and should vary between the maximum and minimum size of the actual lists. Then, to compose each simulated list, a draw was performed without substitution among the 216 items mentioned in the actual lists, ensuring that the items appeared with random frequency and order. Finally, the Saliency Index of each item was calculated according to the formula $S = (\sum((L - R_j + 1)/L))/N$, where “L” is the length of a list, R_j is the rank of item j in the list and N is the number of lists in the sample, like proposed by Smith & Borgatti (1997). The procedure was repeated 300 times, creating a null scenario with 45,900 randomly generated saliency values (Fig. 1).

After establishing the null scenario, we calculated the actual saliency values of the items cited in the free lists (Tab.1 - Column 2), and then we calculated the probability of occurrence (p-value) of the actual saliency value of each item within the null scenario created (Tab. 1 - Column 3). The saliency values calculated using the Smith & Borgatti Index (1997) can vary between “0.0” (items with extremely low saliency) and “1.0” (items with extremely high saliency), depending on the frequency and position of each item in the analyzed list set. In order to verify the influence of each one of these variables in the composition of the saliency index, we calculated the frequency (Tab. 1 - Column 4) and the average position (Tab. 1 - Column 6) of each cited item.

We then used the data from the simulated populations to create a null scenario for each variable. From that, we verified whether the values of frequency and mean position are different than expected at random (Tab. 1 - Column 5 and 7).

The calculated values for our data ranged from 0.4805 (“arruda”) to 0.0001 (“azeitona preta”). Only the 38 items with the highest saliency values (0.4805 – 0.0487) had p-values low enough to be considered different from randomly generated values and, therefore, can be considered as salient. The 42 subsequent items showed saliency values between 0.0729 and 0.0266 and the calculated p-values are not significant. Finally, the 136 items with lower saliency values (0.0266 – 0.0001) also presented values of $p < 0.05$. The item that presented the highest frequency of citation (barbatimão) was cited in 111 lists and 52 items were cited only once. Citation frequency values greater than 20 or lower than 9 were considered statistically significant. Regarding the average position, only items with a mean ranking above 9 or below 20 presented statistically significant values.

Although the identification of breaks in the sequence of saliency values or in a scree plot of the frequency of the cited items (Quinlan 2005; Borgatti & Halgin 2013) is the least arbitrary method for the analysis of free lists, given its subjectivity, it does not allow us to unambiguously decide the boundaries between the salient items. When analyzing our data through the method of identifying breaks in saliency values, for example, we can consider that the first break occurs between the saliency values of “hortelã” (0.3976) and “jatobá” (0.2796) (Tab. 1 - Column 2), since the distance between the two saliency values is about 0.1, representing a drop of almost 30% in the sequence of values. When we plot our frequency values on a scree-plot (Fig. 2), we find a first break between the “janaguba” (104 citations) and the “malva do reino” (95 citations). The second observed break is slightly larger than the first and coincides with the break identified in the sequence of saliency values (hortelã - 89 citations and jatobá - 73 citations). The results of the analyzes performed with the method we propose suggest that, in addition to

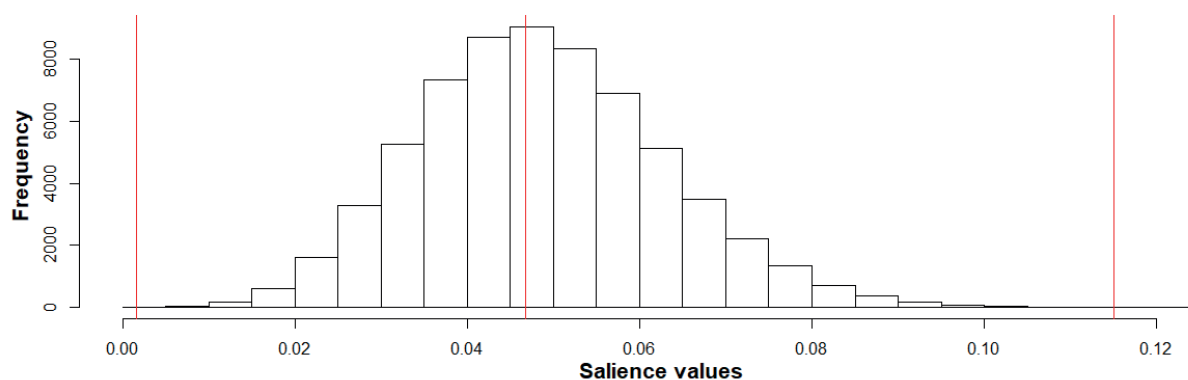


Figure 1. Distribution of the saliency values calculated from lists of simulated populations with the same characteristics of 153 free-lists of medicinal plants in a human population living near FLONA Araripe (Brazil). The vertical lines highlight the actual saliency values calculated three plants cited in the free lists with the highest, lowest and average saliency values.

Table 1. Items cited in 153 free-lists about medicinal plant used by a human community living near FLONA Araripe (Brazil), followed by their respective values of salience and p-values. The salience index was calculated according to the formula proposed by Smith & Borgatti (1997). The p-values of salience show the probability that the salient values occur in a null scenario, calculated from simulated populations with similar characteristics to the real one, using Monte Carlo techniques. The Mean position and Frequency of each item in the lists and their respective p-values are also shown. The values marked with asterisk presented statistically significant values ($p < 0.05$).

Item	Salience	Salience (p.value)	Frequency	Frequency (p.value)	Mean position	Mean position (p.value)
arruda	0.4805	0.0000*	105	0.0000*	7.2857	0.0066*
barbatimão	0.4526	0.0000*	111	0.0000*	8.8739	0.0311*
janaguba	0.4483	0.0000*	104	0.0000*	8.6442	0.0258*
malva do reino	0.4410	0.0000*	95	0.0000*	7.2842	0.0066*
alecrim	0.4397	0.0000*	94	0.0000*	7.2766	0.0066*
hortelã	0.3976	0.0000*	89	0.0000*	8.191	0.0173*
jatobá	0.2796	0.0000*	73	0.0000*	11.0959	0.1485
malva da costa	0.2484	0.0000*	63	0.0000*	9.4603	0.0509
cidreira	0.244	0.0000*	63	0.0000*	10.6349	0.1133
mangaba	0.2359	0.0000*	59	0.0000*	10.4915	0.1031
ameixa	0.2345	0.0000*	65	0.0000*	11.2154	0.1582
alcansus	0.2211	0.0000*	59	0.0000*	13.3559	0.3966
eucalipto	0.2197	0.0000*	68	0.0000*	14.1912	0.5092
pequi	0.2006	0.0000*	59	0.0000*	12.1356	0.2493
capim santo	0.1912	0.0000*	54	0.0000*	11.8148	0.2154
velame	0.175	0.0000*	51	0.0000*	12.9804	0.3475
aroeira	0.1682	0.0000*	50	0.0000*	12.92	0.3415
podoia	0.1533	0.0000*	37	0.0000*	11.8108	0.215
cajueiro	0.1465	0.0000*	46	0.0000*	12.3043	0.2679
mentruz	0.1381	0.0000*	42	0.0000*	14.881	0.4026
marcela	0.1297	0.0000*	42	0.0000*	13.381	0.4005
quebra faca	0.1151	0.0000*	42	0.0000*	17.619	0.1386
laranja	0.1108	0.0001*	37	0.0000*	13.2703	0.3858
orelha de onça	0.1091	0.0002*	32	0.0000*	15.3125	0.3499
anador	0.1062	0.0002*	26	0.0017*	8.2692	0.0183*
sicupira	0.1042	0.0003*	27	0.0008*	14.037	0.4887
novalgina	0.1025	0.0004*	26	0.0017*	8.1923	0.0173*
papaconha	0.1011	0.0006*	34	0.0000*	13.6471	0.4365
babosa	0.0988	0.0009*	30	0.0000*	15.8	0.2961
boldo	0.0976	0.0012*	25	0.0036*	12.44	0.2846
faveira	0.0951	0.0018*	33	0.0000*	12.0909	0.2438
catuaba	0.0929	0.0026*	33	0.0000*	16.697	0.2081
quina-quina	0.091	0.0034*	36	0.0000*	20.5	0.0304*
romã	0.0895	0.0041*	31	0.0000*	17.129	0.1726
araçá	0.088	0.0054*	25	0.0036*	10.88	0.1314
lorma	0.086	0.0077*	20	0.073	10.1	0.0808
imburana	0.0788	0.0232*	26	0.0017*	17.9231	0.1197
vassourinha	0.0732	0.0487*	25	0.0036*	17.96	0.1179
ariticum	0.0729	0.0509	18	0.1738	12.4444	0.2846
sete dores	0.0718	0.0591	19	0.1157	12.6316	0.307
alfavaca	0.0708	0.0665	21	0.0437*	17.0952	0.175
ciriguela	0.0698	0.0749	19	0.1157	12.4211	0.2825
crista de galo	0.0696	0.0765	24	0.0072*	17.6667	0.1351
girassol	0.0695	0.0773	25	0.0036*	21.12	0.0204*
malva branca	0.0681	0.0902	22	0.0252*	18.9091	0.0732
quixaba	0.0652	0.124	26	0.0017*	20.7308	0.0262*
mostarda	0.0639	0.1417	20	0.073	20.6	0.0284*
abacate	0.0596	0.2116	24	0.0072*	22.625	0.0081*
gonçalave	0.0585	0.2339	20	0.073	16.6	0.2174
pau pra tudo	0.0584	0.2364	15	0.4461	13.2	0.3783
erva doce	0.0573	0.2592	18	0.1738	14.6111	0.4353
pau ferro	0.0563	0.2807	17	0.2481	17.4118	0.1525



Table 1. Cont.

Item	Saliency	Saliency (p.value)	Frequency	Frequency (p.value)	Mean position	Mean position (p.value)
noz moscada	0.0557	0.2946	26	0.0017*	22.6538	0.008*
quebra pedra	0.0549	0.3133	21	0.0437*	18.2857	0.101
limão	0.052	0.3856	23	0.0137*	18.3478	0.0975
maracujá do mato	0.0468	0.4727	20	0.073	21.9	0.0129*
endro	0.0464	0.461	16	0.3409	17.6875	0.1348
jasmim laranja	0.0447	0.4147	15	0.4461	16.9333	0.1874
espinho cigano	0.0441	0.3971	16	0.3409	16.0625	0.2688
manjeriço	0.044	0.3944	15	0.4461	15.3333	0.3487
angico	0.0404	0.2964	14	0.5539	18.4286	0.0928
alho	0.0403	0.2938	18	0.1738	20.1667	0.0362*
cana	0.0396	0.2756	18	0.1738	22.1667	0.0108*
congonha	0.0396	0.2756	16	0.3409	21.6875	0.0147*
colônia	0.0371	0.2162	13	0.4448	16	0.2781
carrapicho de boi	0.036	0.193	12	0.3346	19.25	0.0622
pruma	0.0345	0.1629	8	0.0499*	9.625	0.058
batata de purga	0.0342	0.1568	13	0.4448	20.8462	0.0243*
goiabinha	0.0329	0.1324	9	0.0919	17.5556	0.1424
mamona	0.0329	0.1324	14	0.5539	18.7857	0.0783
cebola branca	0.0317	0.113	13	0.4448	19.6923	0.048*
xanana	0.0304	0.0934	9	0.0919	15.4444	0.335
mamão	0.0301	0.0889	11	0.2365	20.4545	0.0309*
andu	0.0299	0.0863	13	0.4448	20.6154	0.0279*
gingibre	0.0296	0.0825	12	0.3346	22.5	0.009*
imbiriba	0.029	0.0756	9	0.0919	19.4444	0.0552
japicanga	0.0289	0.0743	9	0.0919	19	0.0718
manjerioba	0.0286	0.0711	7	0.0234*	12.4286	0.2838
goiaba branca	0.0274	0.0574	8	0.0499*	16.625	0.2148
pau piranha	0.0266	0.0502	8	0.0499*	16.375	0.2371
pinhão roxo	0.0265	0.0492	8	0.0499*	16.75	0.2043
maracujá	0.0258	0.0429	9	0.0919*	16.5556	0.2206
dipirona	0.0255	0.0405	5	0.0036*	5	0.0004*
café	0.0254	0.0397	8	0.0499*	16.625	0.2148
goiaba	0.0254	0.0397	8	0.0499*	17	0.1855
jarrinha	0.0251	0.0377	6	0.0101*	16.5	0.2273
marmeleiro	0.0247	0.0348	9	0.0919	22.8889	0.0068*
linhaça	0.024	0.0304	10	0.1545	19.5	0.0541
catingueira	0.0239	0.0299	6	0.0101*	14.1667	0.5068
mororó	0.0234	0.0269	9	0.0919	21	0.0226*
pata de vaca	0.0233	0.0263	8	0.0499*	23	0.0065*
caninana	0.0232	0.0258	6	0.0101*	12.3333	0.2702
urucum	0.0229	0.0241	9	0.0919	23.6667	0.004*
salsa	0.0225	0.0221	6	0.0101*	15.3333	0.3487
juazeiro	0.0215	0.0175	9	0.0919	16.1111	0.2626
tamburi	0.0201	0.0127	7	0.0234*	16.1429	0.2586
cravo	0.02	0.0125	8	0.0499*	24.125	0.0027*
melancia da praia	0.0191	0.0102	6	0.0101*	21.3333	0.0182*
mentrasto	0.0191	0.0102	6	0.0101*	19	0.0718
jaborandi	0.019	0.0099	9	0.0919	30.4444	0.0000*
carrapicho de agulha	0.0188	0.0094	6	0.0101*	14.1667	0.5068
batata de tiú	0.0183	0.0084	9	0.0919	29.2222	0.0001*
maconha	0.0176	0.0067	6	0.0101*	15.5	0.3303
pitanga	0.0162	0.0044	5	0.0036*	17.2	0.168
chumbim	0.0157	0.0037	4	0.0009*	11.75	0.2095
alecrim do mato	0.0156	0.0036	5	0.0036*	12.6	0.3042
gericó	0.0156	0.0036	4	0.0009*	15.25	0.3581
muçambê	0.0156	0.0036	6	0.0101*	19.6667	0.0481*



Table 1. Cont.

Item	Salience	Salience (p.value)	Frequency	Frequency (p.value)	Mean position	Mean position (p.value)
murici	0.0153	0.0033	6	0.0101*	19.8333	0.0444*
acerola	0.0151	0.0031	6	0.0101*	18.8333	0.0764
podaico	0.0151	0.0031	5	0.0036*	26.6	0.0004*
bom nome	0.0149	0.0029	7	0.0234*	24.5714	0.0019*
salsa parrilha	0.0148	0.0028	3	0.0002*	9	0.0361*
pega pinto	0.0142	0.0024	6	0.0101*	22.8333	0.0072
açafrao	0.014	0.0022	5	0.0036*	18.2	0.1051
pinhao bravo	0.0129	0.0015	3	0.0002*	14	0.4887
canela	0.0127	0.0014	4	0.0009*	17	0.1855
carmelitana	0.0125	0.0013	3	0.0002*	17	0.1855
camomila	0.0118	0.0011	7	0.0234*	25.8571	0.0007*
jurubeba	0.0118	0.0011	6	0.0101*	24.5	0.0021*
abacaxi	0.0115	0.001	7	0.0234*	18.8571	0.0754
melancia	0.0112	0.0008	4	0.0009*	27.25	0.0003*
banana	0.0111	0.0007	4	0.0009*	22	0.0125*
milindro	0.0109	0.0007	2	0.0001*	7.5	0.0089*
chuchu	0.0107	0.0006	7	0.0234*	28	0.0002*
catolé	0.0105	0.0006	3	0.0002*	18.3333	0.0987
coco	0.0105	0.0006	3	0.0002*	19.3333	0.0591
vamora	0.0104	0.0006	2	0.0001*	7.5	0.0089*
contra erva	0.0101	0.0004	4	0.0009*	25	0.0013*
catuaba cipó	0.0098	0.0004	4	0.0009*	20.25	0.0347*
mandioca	0.0098	0.0004	4	0.0009*	19	0.0718
tipi	0.0098	0.0004	4	0.0009*	25.5	0.001*
genipapim	0.0096	0.0004	3	0.0002*	13	0.3571
sabugueira	0.0095	0.0003	2	0.0001*	11	0.1434
laranja da terra	0.0094	0.0003	4	0.0009*	18	0.1179
agrião	0.0092	0.0002	4	0.0009*	28.25	0.0001*
bonina	0.0092	0.0002	4	0.0009*	29.75	0.0000*
mangueira	0.0091	0.0002	3	0.0002*	14.3333	0.4739
rosa doce	0.0087	0.0002	2	0.0001*	13.5	0.4193
tore	0.0086	0.0002	2	0.0001*	11.5	0.1854
cabacinho	0.0083	0.0002	2	0.0001*	16	0.2781
rabo de reposa	0.0082	0.0002	3	0.0002*	16.3333	0.2415
visgueiro	0.0081	0.0002	3	0.0002*	23.6667	0.004*
betonca	0.0076	0.0001	7	0.0234*	19.8571	0.0435*
gergelim	0.0074	0.0001	4	0.0009*	24.5	0.0021*
pinha	0.007	0.0001	2	0.0001*	27	0.0004*
neem	0.0068	0.0001	2	0.0001*	24.5	0.0021*
pitomba	0.0068	0.0001	2	0.0001*	18.5	0.0908
melão	0.0066	0.0001	3	0.0002*	25.3333	0.0011*
rosamelia	0.0062	0.0001	1	0.0000*	3	0.0000*
pau cardoso	0.0061	0.0000	1	0.0000*	4	0.0001*
vick	0.0058	0.0000	1	0.0000*	8	0.015*
quiabo	0.0056	0.0000	1	0.0000*	7	0.005*
balsamo	0.0054	0.0000	3	0.0002*	25	0.0013*
citronela	0.0054	0.0000	1	0.0000*	6	0.0015*
favela	0.0053	0.0000	1	0.0000*	10	0.0773
palma	0.005	0.0000	2	0.0001*	19	0.0718
malva	0.0049	0.0000	2	0.0001*	31.5	0.0000*
pau de leite	0.0049	0.0000	1	0.0000*	13	0.3571
mucuná	0.0047	0.0000	1	0.0000*	5	0.0004*
alento	0.0046	0.0000	1	0.0000*	12	0.2386
jurema	0.0046	0.0000	2	0.0001*	23.5	0.0047*
malícia	0.0046	0.0000	1	0.0000*	4	0.0001*
algodão	0.0045	0.0000	2	0.0001*	25.5	0.001*



Table 1. Cont.

Item	Saliency	Saliency (p.value)	Frequency	Frequency (p.value)	Mean position	Mean position (p.value)
boa noite	0.0045	0.0000	1	0.0000*	15	0.3922
lima	0.0044	0.0000	1	0.0000*	14	0.4887
cipó de vaqueiro	0.0043	0.0000	1	0.0000*	10	0.0773
cassatinga	0.0042	0.0000	1	0.0000*	22	0.0125*
alho roxo	0.004	0.0000	2	0.0001*	27	0.0004*
bambu	0.004	0.0000	1	0.0000*	10	0.0773
chapéu de couro	0.004	0.0000	1	0.0000*	10	0.0773
capim de planta	0.0039	0.0000	2	0.0001*	28	0.0002*
chucalinho	0.0039	0.0000	1	0.0000*	14	0.4887
zezinho	0.0039	0.0000	1	0.0000*	16	0.2781
abobora	0.0038	0.0000	1	0.0000*	18	0.1179
cedro	0.0038	0.0000	1	0.0000*	12	0.2386
fumo	0.0038	0.0000	2	0.0001*	28.5	0.0001*
caninana roxa	0.0037	0.0000	1	0.0000*	22	0.0125*
beterraba	0.0036	0.0000	2	0.0001*	37	0.0000*
eucalipim	0.0036	0.0000	1	0.0000*	11	0.1434
jerimum	0.0035	0.0000	1	0.0000*	20	0.0408*
cana de macaco	0.0034	0.0000	1	0.0000*	12	0.2386
coentro	0.0033	0.0000	1	0.0000*	32	0.0000*
macaxeira	0.0033	0.0000	1	0.0000*	21	0.0226*
milona	0.0032	0.0000	1	0.0000*	20	0.0408*
pimenta	0.0032	0.0000	2	0.0001*	22	0.0125*
batata doce	0.0031	0.0000	1	0.0000*	10	0.0773
braúna	0.0031	0.0000	1	0.0000*	23	0.0065*
ipim roxo	0.0031	0.0000	1	0.0000*	33	0.0000*
babaçu	0.0025	0.0000	1	0.0000*	20	0.0408*
macaúba	0.0023	0.0000	2	0.0001*	31	0.0000*
feijão de corda	0.0022	0.0000	3	0.0002*	31	0.0000*
mandacaru	0.002	0.0000	1	0.0000*	22	0.0125*
pau pereiro	0.002	0.0000	1	0.0000*	10	0.0773
cambuim	0.0019	0.0000	1	0.0000*	28	0.0002*
língua de vaca	0.0019	0.0000	1	0.0000*	18	0.1179
Maracujá peroba	0.0019	0.0000	1	0.0000*	26	0.0006*
salsa caroba	0.0017	0.0000	1	0.0000*	37	0.0000*
imburana de cambao	0.0016	0.0000	1	0.0000*	46	0.0000*
maracujá de estralo	0.0016	0.0000	1	0.0000*	35	0.0000*
cebola vermelha	0.0015	0.0000	2	0.0001*	32.5	0.0000*
cominho	0.0014	0.0000	1	0.0000*	35	0.0000*
Imbu	0.0014	0.0000	1	0.0000*	37	0.0000*
algaroba	0.0012	0.0000	1	0.0000*	18	0.1179
cenoura	0.0012	0.0000	1	0.0000*	32	0.0000*
pimenta do reino	0.0012	0.0000	1	0.0000*	36	0.0000*
fruta bola	0.0011	0.0000	2	0.0001*	28.5	0.0001*
pichuris	0.0011	0.0000	1	0.0000*	51	0.0000*
cidreira brava	0.0009	0.0000	1	0.0000*	21	0.0226*
Maracujá do mato	0.0009	0.0000	1	0.0000*	34	0.0000*
mamelada	0.0007	0.0000	1	0.0000*	42	0.0000*
milho	0.0007	0.0000	2	0.0001*	33	0.0000*
tingui	0.0006	0.0000	1	0.0000*	29	0.0001*
coroa de frade	0.0005	0.0000	1	0.0000*	36	0.00000*
sicupira do Amazonas	0.0005	0.0000	1	0.0000*	14	0.4887
azeitona preta	0.0001	0.0000	1	0.0000*	60	0.00000*



the difficulty in identifying the boundaries between the salient items, the breaks identification method (Quinlan 2005; Borgatti & Halgin 2013) underestimates the quantity of items that should be considered as salient. Through this procedure, only the six items with higher salience values would actually be considered salient. That is about six times less than the amount of items identified with the method we propose.

Since the scree-plot consists of drawing a graph with only the frequency values of the items cited in the lists, implicitly it assumes that this variable is sufficient to define the salience. However, by decomposing the salience value, we find that an item need not necessarily be more frequent than expected at random to present a statistically significant salience value, such as the “lorma”, in our data set (Tab. 1). Likewise, an item may present a citation frequency higher than that expected at random and still not be among the items with statistically significant salience values, such as “quebra-pedra” and “Lemon” (Tab. 1), because it does not occur among the highest-ranking positions. Finally, our data also showed that frequency and mean position may act in opposite directions in the composition of the salience index. This is the case of “quixaba” and “avocado”, which present a statistically significant mean position although they present a frequency of citation higher than expected at random. However, it occurs because they occupy very low positions in the lists.

Discussion

The significantly higher salience values shows that the importance that the interviewed individuals attribute to these items privileges themselves at the moment of

the construction of the free list. This results in a higher frequency of citation and/or an average position than expected by random reasons.

The understanding that, among the elements recorded through free lists only some items with higher salience values can be really salient, has already been emphasized in the literature for a while (Borgatti & Halgin 2013; Quinlan 2005). The choice of these items is often made in an arbitrary way. Tol *et al.* (2018), for example, when analyzing free lists on maternal health problems, explain that items with a Salience Index close to “1” are “high salience values”. Nevertheless, the five most salient items in the free lists ($S = 0.51$ to 0.21) continue to be discussed without any justification for the threshold. Similarly, Wong *et al.* (2015) emphasized that high salience values indicate items of high importance, however, they have not defined how high these values should be. After analyzing the salience values, the authors highlight one to three items with higher salience values (in each category considered) among those cited in a study on the perception of potential clients about the advantages and disadvantages of acquiring health insurance.

The 42 elements whose salience index is not statistically significant when compared to a random distribution generated represent the items on the lists are cited by respondents without receiving any particular prominence. These items may correspond to the content responsible for the heterogeneity of knowledge existing in human populations. This result corroborates the pattern found in studies with medicinal plants, for example, which demonstrate that only a small group of species is known to most people (Ferreira Júnior & Albuquerque 2015). These items would be represented in the analysis suggested here with higher values of and $p < 0.05$.

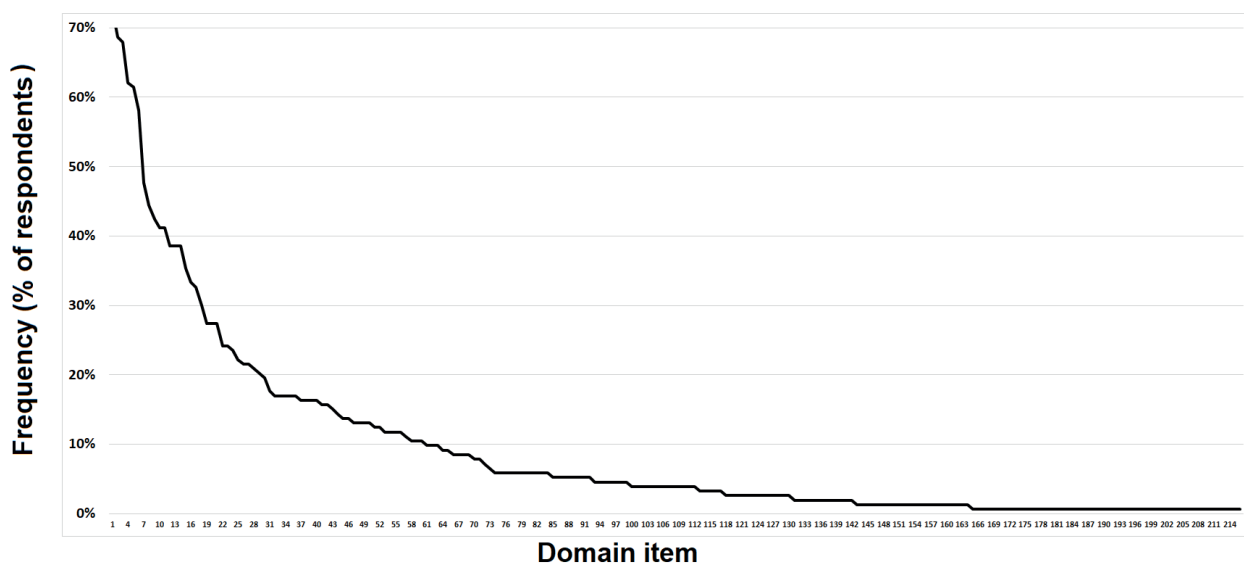


Figure 2. Scree-plot showing the percentual of citation of each item from 153 free-lists about medicinal plant used by a human community living in FLONA Araripe (Brazil).



However, it should be noted that knowledge about medicinal plants is dynamic and subject to intracultural variation, related to factors such as age, gender, income, social roles played locally, among others (Almeida *et al.* 2012; Quinlan e Quinlan 2007; Hanazaki *et al.* 2000; Torres-Avilez *et al.* 2016). An alternative explanation for the items with non- statistically significant salience values is that these items may represent this socio-cultural heterogeneity, reflecting the most relevant sets of plants for people in different intracultural contexts. The analysis of free lists of different socio-cultural groups (gender, age, income, work specialization, etc) of people belonging to the same human population can be performed to verify if the subgroups present different items with statistically significant salience values, which would reinforce this hypothesis.

The ethnobiological studies that evaluate intracultural variation of knowledge tend to use only the number of items cited by each social group evaluated as the main descriptor for this variation. This methodology is quite frequent in studies related to the effect of gender (Torres-Avilez & Albuquerque 2017). The comparison of the salient items ($p < 0.05$) of the whole community with those of the social groups may favor the more comprehensively understanding of the local differences of a given cultural domain, by indicating which items are characteristic of each group.

136 of the 216 items present in the lists that we use can be considered idiosyncrasies, by the same criteria adopted in our analysis. These items presented unusually low salience values, with very low probability (p -value) of being produced on randomly generated lists. Such items are known by very few people or cited at last in the free lists, and for these reasons have been interpreted by the literature as items of little or no cultural importance (Borgatti & Halgin 2013) or mistakes (Quinlan 2005).

The existence of the need to identify idiosyncrasy among the items quoted in free lists has already been discussed by Borgatti & Halgin (2013). The authors suggest that the items cited by only one informant should be discarded. However, in addition to recognizing that this is not sufficient to eliminate all necessary items, this criterion is based on only one aspect of the salience, the frequency of citation.

The exclusion of items described in these studies, besides totally arbitrary, does not allow the understanding of the factors that may be leading to these idiosyncrasies. Depending on the structuring of the local medical system, for example, knowledge about the use of some medicinal plants may be concentrated in a few people who perform very specific functions, such as local experts. This would make plants of high cultural importance have very low salience values. Cultural information is also subject to “errors” during the process of social transmission (Laland & Brown 2002), a factor that can contribute to the appearance of these little transmitted local items. In addition, human beings have the capacity to create knowledge, to innovate, usually as an adaptive response to environmental situations (Boyd

et al. 2011), a factor that can also generate idiosyncratic information.

Moreover, the sharing of cultural information in human populations is subject to temporal changes, that is, information that has been very frequent in the past may become infrequent in the present time depending on the environmental situation experienced. However, the opposite is also true (Mesoudi 2011). For example, extreme drought events in caatinga areas in Brazil lead people to use a set of emergency food plants that are resistant to drought and require more complex preparation (Nascimento *et al.* 2012). It is likely that, after a long period of scarcity, only older people will cite most of these emergency plants. Thus, the cultural salience of these emergency food plants will vary depending on the current environmental situation and the constancy of extreme drought events. Thus, comprises the characteristics of the least salient items and how they are distributed across different socio-cultural groups in the community may allow the understanding of the dynamics of entry and exit of cultural information in human populations.

Final considerations

In our calculations, we applied the formula proposed by Smith & Borgatti (1997) to calculate the Salience Index, because this index is the most used in ethnobiological studies. However, the analysis we propose is applicable to the interpretation of the salience calculated by any other formula (e.g. Sutrop 2001; Robbins *et al.* 2017) or even for free-recall studies.

The use of simulated populations to generate a null model and subsequent verification of statistical significance of the salience values opens up new perspectives for the studies that use the free list as a technique for collecting data. In addition to objectively indicating which items are more prominent in relation to the others as to their frequency and position in the lists, the present analysis also shows which items have salience values significantly lower than expected by chance. This could be a result of the production of individual knowledge (innovations), recent information inputs (immigration), changes of the original information (mutations) or low mnemonic relevance.

The present analysis is, therefore, a useful tool for the understanding of processes related to cultural relevance and/or mnemonic of items, thus contributing to the test of future hypotheses in different areas of social sciences and ethnobiology.

Acknowledgements

We acknowledge the CNPq for the research productivity grant awarded to UPA and post-doctoral fellowship awarded to ALBN (Number: 151518/2018-1). Contribution of the INCT Ethnobiology, Bioprospecting and Nature Conservation, certified by CNPq, with financial support



from FACEPE (Foundation for Support to Science and Technology of the State of Pernambuco - Grant number: APQ-0562-2.01/17).

References

- Almeida CFCBR, Ramos MA, Silva RRV, *et al.* 2012. Intracultural variation in the knowledge of medicinal plants in an urban-rural community in the atlantic forest from Northeastern Brazil. *Evidence-Based Complementary and Alternative Medicine* 2012: 1-15.
- Borgatti S. 1990. Using ANTHROPAC To Investigate a cultural domain. *CAM Newsletter* 2: 8-8.
- Borgatti SP. 1996. ANTHROPAC 4.0. Columbia, Analytic Technologies.
- Borgatti SP, Everett MG, Freeman LC. 2002. Ucinet for Windows: Software for social network analysis. Harvard, Analytic Technologies.
- Borgatti SP, Halgin DS. 2013. Elicitation techniques for cultural domain analysis. In: Schensul JJ, Lecompte MD. (eds.) *Specialized ethnographic methods: A mixed methods approach*. Lanham, AltaMira Press. p. 80-116.
- Boyd R, Richerson PJ, Henrich J. 2011. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 108: 10918-10925.
- Brewer DD. 1985 Cognitive indicators of knowledge in semantic domains. *Pile Sort Analysis of siSwati Terms for Acute Respiratory Infections* 5: 107-128.
- Cartaxo SL, de Almeida Souza MM, de Albuquerque UP. 2010. Medicinal plants with bioprospecting potential used in semi-arid northeastern Brazil. *Journal of Ethnopharmacology* 131: 326-342.
- Ferreira Júnior WS, Albuquerque UP. 2015. "Consensus within diversity": An evolutionary perspective on local medical systems. *Biological Theory* 10: 363-368.
- Hanazaki N, Tamashiro JY, Leitao-Filho HF, Begossi A. 2000. Diversity of plant uses in two Caiçara communities from the Atlantic Forest coast, Brazil. *Biodiversity and Conservation* 9: 597-615.
- Ladio A, Lozada M, Weigandt M. 2007. Comparison of traditional wild plant knowledge between aboriginal communities inhabiting arid and forest environments in Patagonia, Argentina. *Journal of Arid Environments* 69: 695-715.
- Laland KN, Brown GR. 2002. *Sense and Nonsense: Evolutionary Perspectives on Human Behaviour*. 1st. edn. Oxford, Oxford University Press.
- Leitão F, Leitão SG, Almeida MZ, Cantos J, Coelho T, Silva PEA. 2013. Medicinal plants from open-air markets in the State of Rio de Janeiro, Brazil as a potential source of new antimycobacterial agents. *Journal of Ethnopharmacology* 149: 513-521.
- Lozano A, Araújo EL, Medeiros MF, Albuquerque UP. 2014. The apparency hypothesis applied to a local pharmacopoeia in the Brazilian northeast. *Journal of Ethnobiology and Ethnomedicine* 10: 1-171.
- Mesoudi A. 2011. *Cultural evolution: How darwinian theory can explain human culture and synthesize the social sciences*. 1st. edn. Chicago, The University of Chicago Press.
- Nascimento ALB, Lozano A, Melo JG, Alves RRN, Albuquerque UP. 2016. Functional aspects of the use of plants and animals in local medical systems and their implications for resilience. *Journal of Ethnopharmacology* 194: 348-357.
- Nascimento VT, Vasconcelos MAS, Maciel MIS, Albuquerque UP. 2012. Famine foods of Brazil's seasonal dry forests: Ethnobotanical and nutritional aspects. *Economic Botany* 66: 22-34.
- Nunes AT, Cabral DLV, Amorim ELC, Santos MVE, Albuquerque UP. 2016. Plants used to feed ruminants in semi-arid Brazil: A study of nutritional composition guided by local ecological knowledge. *Journal of Arid Environments* 135: 96-103.
- Pennec F, Wencelius J, Garine E, Raimond C, Bohbot H. 2012. FLAME v1.0: Free-list analysis under Microsoft Excel. Paris, CNRS.
- Purzycki BG, Jamieson-Lane A. 2017. AnthroTools: An R Package for Cross-cultural ethnographic data analysis. *Cross-Cultural Research* 51: 51-74.
- Quinlan M. 2005. Considerations for collecting freelists in the field: Examples from ethobotany. *Field Methods* 17: 219-234.
- Quinlan MB, Quinlan RJ. 2007. Modernization and medicinal plant knowledge in a caribbean horticultural village. *Medical Anthropology Quarterly* 21: 169-192.
- R Core Team. 2017. R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing. <https://www.R-project.org/>
- Robbins MC, Nolan JM, Chen D. 2017. An improved measure of cognitive salience in free listing tasks. *Field Methods* 29: 395-403
- Robert C, Casella G. 2010. *Introducing Monte Carlo methods with R*. New York, Springer New York.
- Rommey AK, D'Andrade RG. 1964. Cognitive aspects of english kin terms. *American Anthropologist* 66: 146-170.
- Santoro FR, Ferreira Júnior WS, Araújo TAS, Ladio AH, Albuquerque UP. 2015. Does plant species richness guarantee the resilience of local medical systems? A perspective from utilitarian redundancy. *PLOS ONE* 10:e0119826. doi: 10.1371/journal.pone.0119826
- Smith J. 1993. Using ANTHOPAC 3.5 and a spreadsheet to compute a free-list Salience Index. *Cultural Anthropology Methods* 5: 1-3.
- Smith JJ, Borgatti SP. 1997. Salience counts and so does accuracy: Correcting and updating a measure for free-list-item salience. *Journal of Linguistic Anthropology* 7: 208-209.
- Smith JJ, Furbee L, Maynard K, Quick S, Ross L. 1995. Salience counts: A domain analysis of english color terms. *Journal of Linguistic Anthropology* 5: 203-216.
- Sousa DCP, Soldati GT, Monteiro JM, *et al.* 2016. Information retrieval during free listing is biased by memory: Evidence from medicinal plants. *PLOS ONE* 11:e0165838. doi: 10.1371/journal.pone.0165838
- Sutrop U. 2001. List task and a cognitive Salience Index. *Field Methods* 13: 263-276.
- Thompson EC, Juan Z. 2006. Comparative cultural salience: Measures using free-list data. *Field Methods* 18: 398-412.
- Tol WA, Ebrecht BO, Aiyo R, *et al.* 2018. Maternal mental health priorities, help-seeking behaviors, and resources in post-conflict settings: A qualitative study in eastern Uganda. *BMC Psychiatry* 18: 1-15.
- TorresAvilez WM, Albuquerque UP. 2017. Dynamics of social-ecological systems: gender influence in local medical systems. *Ethnobiology and Conservation* 6: 1-6.
- Torres-Avilez WM, Medeiros PM, Albuquerque UP. 2016. Effect of gender on the knowledge of medicinal plants: systematic review and meta-analysis. *Evidence-Based Complementary and Alternative Medicine* 2016: 1-13.
- Weller S, Romney A. 1988. *Systematic data collection*. Vol. 10. Qualitative research methods. California, SAGE Publications.
- Wong CA, Asch DA, Vinoya CM, *et al.* 2015. Seeing health insurance and healthcare.gov through the eyes of young adults. *Journal of Adolescent Health* 57: 137-143.
- Zambrana NYP, Bussmann RW, Hart RE, *et al.* 2018. To list or not to list? The value and detriment of freelisting in ethnobotanical studies. *Nature Plants* 4: 201-204.

