

# 1

## Adjustment of the Ranking of Kernel Words in Light of Cases of Language Contact<sup>1</sup>

---

Baoya CHEN, Zihe LI, Dejiang YU  
Peking University

### Abstract

Ranking of kernel words is an effective means to judge the linguistic genetic relationship. The ranking of kernel words needs to be dynamically adjusted according to the accumulated language materials, so that the difference between the two ranks of core words will be more significant. Based on the data from long-term investigation, we examined three cases of language contact, that is, the contact between Naxi and Chinese, Dai and Chinese, and Chinese dialects and Putonghua. Starting from the actual performance of the kernel words in the language contact, we tried to adjust the ranking of the kernel words, tested the effectiveness of the adjustment, and discussed the methods of ranking adjustment.

---

1. Supported by Major Project of National Social Science Foundation (22&ZD218).

## 1. Introduction

How to distinguish the genetic relationship and contact relationship of languages is a problem that needs to be solved in historical linguistics. It has also been a pressing question that population genetics expects linguists to answer. After years of exploration, academia has realized that language contact is unbounded, and the vertical genetic components and horizontal transmission components are often intertwined; sometimes even the horizontal transmission components of multiple time periods overlap, forming multiple sets of correspondence relationships. This requires distinguishing historical strata, and then determining whether the related words at the earliest stratum are cognates or loanwords.

After working out the related words at the earliest stratum, it is often impossible to use the formal differences to distinguish loanwords from cognates. At present, it is feasible to turn to the basic word ranking of these earliest related words. This method holds that, if the high-rank (most highly stable) basic words take up a larger proportion of the related words, the languages in question are genetically related; if the low-rank (less stable) basic words take up a larger proportion of the related words, these related words are the outcome of a contact relationship (Chen 1996).

However, until now scholars still have differing views on how to determine the degree of stability of words. Different scholars have proposed different basic word sets. How to decide the ranking of basic words is an urgent problem that analysis of the current ranking method is faced with. If the procedure can be worked out, it would be of great significance in determining the genetic relationship of languages.

## 2. Previous Studies

Many scholars believe that there is an absolutely stable basic vocabulary across languages. Swadesh was the first to systematically demonstrate the concept of basic vocabulary, and has proposed two basic vocabularies successively (Swadesh 1952; 1955). However, these two lists are based on a small number of languages. Later, Dolgopolsky's (1986) 23 word list, Jaxontov's 35 word list (Starostin 1991), Lohr's (1998) 61 word list, Holman et al.'s (2008) 40 word list, and so on, all focused on the stability of words in language contact, that is, the basic vocabulary is the most difficult word set to be replaced in language contact. Tadmor et al. (2009) comprehensively considered four factors: borrowing rate, representation rate, analyzability, and age. Based on the statistics of 41 languages, the Leipzig-Jakarta basic vocabulary list was proposed through careful calculation.

Research on language contact revealed that all levels of language can be touched upon in the process of contact (Thomason and Kaufmann 1988). Chen (1996) proposed that language contact is "unbounded but graded"; kernel words can also be borrowed but there are different ranks of borrowability. Therefore, based on Swadesh's (1952; 1955) word list, Chen (1996) puts forward the distinction between the first 100 basic words (high-rank kernel words) and second 100 basic words (low-rank kernel words). The result of testing with published materials shows that the proportion of cognates in high-

rank word sets is higher than that in low-rank word sets among languages with a known genetic relationship (literature materials or other evidence). For example, cognates among Germanic languages, Tai languages, Chinese Mandarin dialects, Dai dialects, Bai dialects, and Yi dialects all follow this distribution. On the contrary, the loanword proportion of high-rank word sets is lower than that of low-rank word sets among languages known to have a contact relationship. For example, the distribution of kernel words between Dai dialects and Southwest Mandarin Chinese, between Chinese and Japanese, and between Chinese and Vietnamese all exemplify this principle. This should not be accidental. There have been discussions on the choice of kernel words, such as those by Chen and Wang (2006), Jiang (2008; 2011), and Huang (2008). Although different word lists have been proposed, the idea of ranking is consistent. Therefore, how to determine the rank of kernel words is the key.

Chen and Li (2012) put forward the basic idea of dynamic ranking adjustment of kernel words, including two algorithm models: (1) The borrowing rate algorithm which automatically searches the database according to the language distribution of each related word in the database of sound correspondence, calculates the borrowing rate of each word automatically, and sorts the words according to the borrowing rate. The more languages in which a word is borrowed, the higher the borrowing rate and the lower the stability. (2) The kernel word ranking adjustment algorithm. According to the borrowing rate calculated by the borrowing rate model, the kernel word ranking adjustment model automatically transfers words with a low borrowing rate to the high-rank word set and words with a high borrowing rate to the low-rank word set. In the process, this model also automatically finds parameters to determine the boundary between high-rank kernel words and low-rank kernel words, so that the average borrowing rate of high-rank kernel words reaches the minimum value and the average borrowing rate of low-rank kernel words reaches the maximum value.

To achieve the goal of dynamic ranking of kernel words, it is important for the database to be based on high-quality materials. It is necessary to conduct a thorough investigation, collect as many morphemes as possible, and then establish rigid sound correspondences among languages. Therefore, at present, we temporarily used materials in the Naxi language and Dai language, which have been investigated on a long-term basis, to conduct a case study on the ranking adjustment of the kernel words, supplemented by the data on Chinese dialects, for which there are plenty of materials.

### **3. Method**

Two cases of ongoing language contact were chosen to observe the process of kernel word replacement. Some of the authors have been conducting long-term surveys of Naxi and Dai (Li 2013; 2014; 2019; 2021; Yu 2020); therefore, it was possible to perform a thorough investigation into the contact between Naxi and Chinese, and the contact between Dai and Chinese. The Jiuhe variety of Naxi and the Qimaba variety of Dai were chosen for the case study.

In each case study, we first determined the kernel vocabulary of Naxi and Dai according to the semantic pairing of Swadesh's (1952; 1955) English version of basic words with the corresponding words in Naxi and Dai, respectively. In order to meet the requirements of empirical research to the greatest extent, we adopted sociolinguistics research methods to observe the process of kernel word replacement from the perspective of "apparent time" (Weinreich, Labov, and Herzog 1968). For each case we collected data on kernel words from multiple speakers, among which young people who were greatly influenced by Chinese accounted for a large proportion. Then, for each kernel word, we counted how many participants had shifted to the corresponding Chinese words, and calculated the replacement rate of each word. Finally, all the kernel words were sorted from low to high according to the replacement ratio. The kernel words originally belonging to the high rank should be relegated to the low rank if they were ranked in the second half of the replacement ratio list. Kernel words originally belonging to the low rank should be promoted to the high rank if they were ranked in the first half.

We also chose several cases of Chinese dialects to illustrate the stability of kernel words. *Hanyu Fangyan Cihui (Vocabularies of Chinese Dialects*, Department of Chinese Language and Literature of Peking University 1995) has collected 1,230 common words from 20 Chinese dialects. The survey was conducted in the mid-1980s, and 1–3 middle-aged or old (40–70 years old) participants were selected for each dialect. In the past two years, we selected four dialects and performed a new investigation with middle-aged participants in the same way, and compared their words with the records in *Hanyu Fangyan Cihui*. In this way, we could obtain the changes in the kernel words of these dialects in the past 40 years. It can be inferred from societal conditions in China that the changes of kernel words in the past 40 years are mainly due to the influence of Putonghua on dialects. However, due to the small number of participants included for each dialect, we could not fully observe the variation. Therefore, we have only taken the change of the kernel words of the dialects as secondary evidence.

## **4. The Case of Jiuhe Naxi**

### **4.1 Background and Research Design**

Jiuhe Bai Autonomous Township is located in southeast Yulong Naxi Autonomous County, Yunnan Province. The township government is about 50 kilometers away from the county proper. The village has 11 villagers' committees and 81 villagers' groups, with a population of about 28,600 (Rural Socio-economic Survey Division of National Bureau of Statistics 2020: 568), of which the Bai ethnic minority accounts for about 52%, the Naxi ethnic minority accounts for about 39%, and other ethnic groups, including the Han, Pumi, Lisu, and Tibetan, account for about 9%.

The Naxi variety spoken by the Naxi people in Jiuhe Bai Autonomous Township belongs to the western dialect of the Naxi language (He and Jiang 1985), which is

mutually intelligible with the Naxi language in Lijiang. The vast majority of Naxi people also use Chinese. People over the age of 40 use Southwest Mandarin more often, and people below that age mainly use Putonghua. If there is a large Bai population in a village committee, Naxi people over age 60 can also speak the Bai language, but most Naxi people under 60 cannot speak the Bai language.

We conducted two surveys on Jiuhe Naxi phonology and vocabulary from July to August 2021 and July to August 2022. Among the participants, there was a male born in 1975 and a female born in 1964 whose kernel words were basically not replaced; therefore, their kernel words were taken as the initial state that has not been touched through language contact.

The division of conceptual space in Naxi vocabulary is quite different from that in English and Chinese, so it is necessary to discuss the kernel word list first.

Compared with Swadesh's basic word list, Jiuhe Naxi has 12 compounds which should be eliminated from the kernel word list:

- Bark /ndzu<sup>21</sup>-γu<sup>33</sup>/: “tree” + “skin”
- Knee /na<sup>33</sup>-kv<sup>21</sup>-tv<sup>55</sup>.lv<sup>33</sup>/: black + “shell” + “a small lump of”
- Dull /mə<sup>33</sup>-tʰa<sup>55</sup>/: NEG + “sharp”
- Fruit /ndzu<sup>21</sup>-kv<sup>33</sup>-ndzu<sup>21</sup>-ly<sup>33</sup>/: “tree” + “to bear” + “tree” + “small round object”
- How /ə<sup>33</sup>.tsɿ<sup>33</sup>-be<sup>33</sup>/: “what” + “do”
- Hunt /çy<sup>21</sup>-ndy<sup>33</sup>/: “beast” + “chase”
- We /ŋx<sup>33</sup>-ŋgu<sup>21</sup>/: “I” + PL<sup>2</sup>
- They /tʰ<sup>33</sup>-ŋgu<sup>21</sup>/: “he/she” + PL
- Breasts /na<sup>55</sup>-po<sup>21</sup>/: “milk” + “bump”
- Here /tʂ<sup>33</sup>-ku<sup>21</sup>/: “this” + “place”
- There /ə<sup>55</sup>.tʂ<sup>33</sup>.ku<sup>21</sup>/: “that” + “place”
- Where /ze<sup>13</sup>-ku<sup>21</sup>/: “what” + “place”

There are three words in Swadesh's list that do not have a corresponding concept in the Naxi language: because, if, sea.

The first two words are grammatical words expressing abstract meaning. The last word “sea” is never used in Naxi because Naxi people have traditionally lived far from the sea.

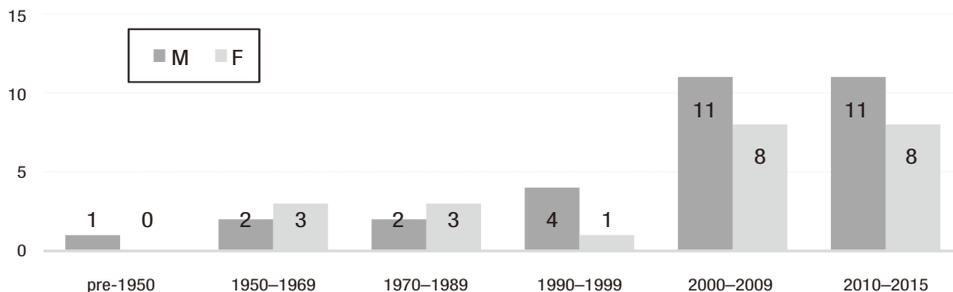
The significance of two words in Swadesh's list has merged with that of other words as unitary concepts in Naxi: /kʰu<sup>33</sup>/ denotes both “leg” and “foot”, and /fv<sup>33</sup>/ is used for both “feather” and “hair”.

In addition, there are 19 words whose meanings are further divided and expressed by multiple words in Naxi. These words include the following:

- 15 words each corresponding to two items: kill, come, give, speak, good, blow, dig, fight, in, old, pull, split, thick, thin, year;
- Three words each corresponding to three items: bite, cut, hold/take;
- One word corresponding to four items: fall.

2. Abbreviations: FUT, future tense; intr., intransitive; NEG, negative; PL, plural; PST, past tense; tr., transitive.

**Figure 1.1 The age and gender distribution of the Jiuhe Naxi speakers**



The specific semantic divisions are illustrated in Table 1.1 and 1.2. Therefore, the kernel word list of Jiuhe Naxi should include 208 items.

The special survey of kernel words was carried out in August 2022<sup>3</sup>, using the Swadesh 200 basic words list compiled by Chen (1996). Data of the 54 participants was collected. The age and gender distribution of the participants are shown in Figure 1.1.

The survey was conducted in Putonghua. In order to reduce errors caused by any misunderstanding in communication, certain contexts were used when asking about nouns. For example, the question for “liver” was “How do you say the liver of a pig in Naxi? (豬的肝用納西話怎麼說?)”; for “feather”, the question was “What do you mean by birds’ feathers in Naxi? (鳥的毛用納西話怎麼說?)”. When asking about verbs and body parts, actions or gestures were also used. When investigating function words, people were asked to try to say a complete sentence containing the word. For example, for “and” they were asked to say “He and I are here (我和他都來了)”.

If the participant said “I don’t know” for a word, or “just use Chinese words”, or could not answer even after thinking for a long time (more than 10 seconds, or a few more seconds for older people), the word was considered as replaced.

## 4.2 Results

The survey data shows that the kernel words of the participants under 40 years old have obviously changed. Most of these participants found that some of the kernel words were difficult to recall. No instances were found where Naxi kernel words had been replaced by the corresponding Bai words.

The proportions of high-rank and low-rank kernel words that were substituted are illustrated in Table 1.1 and Table 1.2, respectively.

3. The fieldwork was carried out by the author with eight undergraduate students majoring in linguistics from Peking University. The students are (in alphabetical order) CHEN Jinlin 陳金琳, CHEN Kaidi 陳凱迪, CHEN Ziru 陳子儒, HUANG Xi 黃熙, YANG Ziyang 楊紫嫣, YU Hang 禹航, ZHENG Chen 鄭辰, and ZHENG Zhiyu 鄭旨好.

Table 1.1 High-rank kernel words in Jiuhe Naxi

#	English	Naxi	Note	Proportion substituted	Rank of substitution	Relegation
1	I	ŋɣ <sup>21</sup>		0.07	29	
2	you	ŋɿ <sup>21</sup>		0.19	79	
4	this	tɕ <sup>h</sup> ɿ <sup>33</sup>		0.20	90	
5	that	ə <sup>55</sup> tɕ <sup>h</sup> ɿ <sup>33</sup>		0.35	134	↓
6	who	ə <sup>33</sup> ne <sup>21</sup>		0.19	80	
7	what	ə <sup>33</sup> tsɿ <sup>33</sup>		0.04	11	
8	not	mɣ <sup>33</sup>		0.04	12	
9	all	qɿ <sup>33</sup> xɣ <sup>33</sup>		0.15	57	
10	many	bu <sup>21</sup>		0.07	30	
11	one	qɿ <sup>33</sup>		0.20	91	
12	two	ni <sup>33</sup>		0.17	67	
13	big	qɿ <sup>21</sup>		0.02	4	
14	long	ɕə <sup>21</sup>		0.17	68	
15	small	ci <sup>55</sup>		0.07	31	
16	woman	mi <sup>55</sup>		0.22	101	
17	man	zo <sup>33</sup>		0.13	52	
18	person	çi <sup>33</sup>		0.07	32	
19	fish	ni <sup>33</sup>		0.04	13	
20	bird	v <sub>i</sub> <sup>55</sup> ze <sup>33</sup>		0.07	33	
21	dog	k <sup>h</sup> uɿ <sup>33</sup> ni <sup>33</sup>		0.04	14	
22	louse	ɕu <sup>33</sup>		0.65	186	↓
23	tree	ndzu <sup>21</sup>		0.19	81	
24	seed	lə <sup>55</sup>		0.35	135	↓
25	leaf	ndzu <sup>21</sup> p <sup>h</sup> iɿ <sup>33</sup>		0.19	82	
26	root	k <sup>h</sup> uɿ <sup>33</sup>		0.52	168	↓
28	skin	ɣu <sup>33</sup> p <sup>h</sup> i <sup>21</sup>		0.50	162	↓
29	flesh	ɕɿ <sup>33</sup>		0.46	157	↓
30	blood	sa <sup>33</sup>		0.06	25	
31	bone	ɕwa <sup>33</sup> lwa <sup>33</sup>		0.17	69	
32	grease	ma <sup>21</sup>		0.50	163	↓
33	egg	(ʔa <sup>21</sup> ) kv <sub>i</sub> <sup>33</sup>		0.02	5	
34	horn	k <sup>h</sup> a <sup>33</sup>		0.50	164	↓
35	tail	ma <sup>33</sup> ta <sup>55</sup>		0.26	112	↓
37	hair	f <sub>v</sub> <sup>33</sup>	(including “feather”)	0.09	40	
38	head	ku <sup>33</sup> ly <sup>33</sup>		0.04	15	

Table 1.1: Continued

#	English	Naxi	Note	Proportion substituted	Rank of substitution	Relegation
39	ear	xe <sup>33</sup> tsv <sup>21</sup>		0.17	70	
40	eye	mjɿ <sup>21</sup> ly <sup>33</sup>		0.09	41	
41	nose	ni <sup>55</sup> mǝ <sup>21</sup>		0.15	58	
42	mouth	nu <sup>55</sup> ta <sup>33</sup>		0.19	83	
43	tooth	hu <sup>33</sup>		0.13	53	
44	tongue	çi <sup>55</sup>		0.17	71	
45	claw	tʂɿ <sup>21</sup>		0.74	199	↓
46	foot	k <sup>h</sup> u <sup>33</sup>	(including “leg”)	0.04	16	
48	hand	la <sup>21</sup>		0.00	1	
49	belly	dv <sup>21</sup> me <sup>33</sup>		0.15	59	
50	neck	ce <sup>33</sup> (pa <sup>21</sup> )		0.24	105	↓
52	heart	nu <sup>55</sup> me <sup>33</sup>		0.20	92	
53	liver	sɿ <sup>55</sup>		0.61	182	↓
54	drink	t <sup>h</sup> u <sup>21</sup>		0.02	6	
55	eat	ndzi <sup>33</sup>		0.00	2	
56	bite	ts <sup>h</sup> a <sup>55</sup>		0.24	106	↓
56	bite	k <sup>h</sup> a <sup>33</sup>	dog bite	0.19	84	
56	bite	t <sup>h</sup> ǝ <sup>55</sup>	mosquito bite	0.28	117	↓
57	see	do <sup>21</sup>		0.22	102	
58	hear	mi <sup>33</sup>		0.09	42	
59	know	ei <sup>33</sup>		0.17	72	
60	sleep	ji <sup>33</sup>		0.02	7	
61	die	ʂɿ <sup>33</sup>		0.07	34	
62	kill	k <sup>h</sup> o <sup>55</sup>	~ an animal	0.17	73	
62	kill	sy <sup>55</sup>	~ people	0.17	74	
63	swim	ji <sup>21</sup> ndzɛ <sup>33</sup>		0.17	75	
64	fly	bi <sup>21</sup>		0.06	26	
65	walk	ji <sup>33</sup>		0.04	17	
66	come	te <sup>h</sup> i <sup>33</sup>	PST	0.04	18	
66	come	lu <sup>33</sup>	FUT	0.15	60	
67	lie	ji <sup>33</sup> (睡覺)		0.20	93	↑
68	sit	ndzi <sup>21</sup>		0.04	19	
69	stand	hy <sup>55</sup>		0.11	47	

#	English	Naxi	Note	Proportion substituted	Rank of substitution	Relegation
70	give	ku <sup>21</sup>	~ back	0.20	94	
70	give	jɿ <sup>55</sup>		0.19	85	
71	speak	ʂɿ <sup>55</sup>	tr.	0.17	76	
71	speak	ku <sup>55</sup> tʂɿ <sup>33</sup>	intr.	0.07	35	
72	sun	ni <sup>33</sup> me <sup>33</sup>		0.39	143	↓
73	moon	xe <sup>33</sup> me <sup>33</sup>		0.41	144	↓
74	star	ku <sup>21</sup>		0.57	175	↓
75	water	ji <sup>21</sup>		0.00	3	
76	rain	hu <sup>21</sup>		0.09	43	
77	stone	lu <sup>33</sup> pa <sup>33</sup>		0.07	36	
78	sand	ʂɿ <sup>55</sup>		0.69	192	↓
79	earth	tʂɿ <sup>33</sup>		0.57	176	↓
80	cloud	ji <sup>21</sup>		0.50	165	↓
81	smoke	mu <sup>55</sup> k <sup>h</sup> u <sup>21</sup>		0.35	136	↓
82	fire	mi <sup>33</sup>		0.07	37	
83	ash	ɣu <sup>55</sup> cy <sup>21</sup>		0.61	183	↓
84	burn	ŋdʒɿ <sup>33</sup>		0.30	121	↓
85	path	zv <sub>i</sub> <sup>33</sup> gv <sub>i</sub> <sup>33</sup>		0.43	150	↓
86	mountain	ɲjɿ <sup>21</sup>		0.22	103	
87	red	hy <sup>21</sup>		0.19	86	
88	green	ha <sup>21</sup>		0.43	151	↓
89	yellow	ʂɿ <sup>21</sup>		0.37	141	↓
90	white	p <sup>h</sup> ə <sup>21</sup>		0.13	54	
91	black	na <sup>21</sup>		0.04	20	
92	night	xo <sup>21</sup> mu <sup>21</sup> k <sup>h</sup> v <sub>i</sub> <sup>55</sup>		0.26	113	↓
93	hot	ts <sup>h</sup> u <sup>33</sup>		0.07	38	
94	cold	c <sup>h</sup> i <sup>55</sup>		0.02	8	
95	full	ʂə <sup>55</sup>		0.15	61	
96	new	ʂɿ <sup>55</sup>		0.26	114	↓
97	good	gv <sub>i</sub> <sup>33</sup>	of personality	0.56	173	↓
97	good	pjɿ <sup>33</sup>	of quality	0.24	107	↓
98	round	we <sup>55</sup> we <sup>33</sup>		0.83	206	↓
99	dry	pv <sub>i</sub> <sup>21</sup>		0.26	115	↓
100	name	mi <sup>21</sup>		0.33	128	↓

**Table 1.2** Low-rank kernel words in Jiuhe Naxi

#	English	Naxi	Note	Proportion substituted	Rank of substitution	Promotion
101	and	ne <sup>33</sup>		0.50	166	
102	animal	cy <sup>33</sup>		0.87	207	
103	back	gu <sup>33</sup> du <sup>33</sup>		0.78	201	
104	bad	k <sup>h</sup> wa <sup>21</sup>		0.20	95	
106	blow	mu <sup>21</sup>	~ the trumpet	0.15	62	↑
106	blow	t <sup>h</sup> u <sup>33</sup>	wind blows	0.44	152	
107	breathe	sa <sup>55</sup>		0.57	177	
108	child	zy <sup>55</sup> zy <sup>33</sup>		0.09	44	↑
109	count	zwa <sup>21</sup>		0.31	123	
110	cut	ha <sup>55</sup>	~ meat	0.19	87	↑
110	cut	ŋga <sup>21</sup>	~ cloth	0.35	137	
110	cut	p <sup>h</sup> jɿ <sup>55</sup> p <sup>h</sup> jɿ <sup>33</sup>	~ (a fish) open	0.89	208	
111	day	ni <sup>21</sup> le <sup>55</sup> gv <sup>33</sup>		0.46	158	
112	dig	ndzi <sup>55</sup>		0.57	178	
112	dig	p <sup>h</sup> a <sup>21</sup>	of animals	0.78	202	
113	dirty	ny <sup>33</sup>		0.81	204	
115	dust	yu <sup>55</sup> ma <sup>55</sup>		0.44	153	
116	fall	mbjɿ <sup>21</sup>	a wall falls	0.31	124	
116	fall	ndo <sup>33</sup>	fall down	0.11	48	↑
116	fall	ty <sup>33</sup>	fall off	0.81	205	
116	fall	gu <sup>33</sup>	rainfall	0.04	21	↑
117	far	k <sup>h</sup> u <sup>33</sup> k <sup>h</sup> wa <sup>33</sup>		0.41	145	
118	father	ɿ <sup>55</sup> ti <sup>33</sup>		0.70	196	
119	fear	zø <sup>33</sup>		0.02	9	↑
120	few	nɿ <sup>33</sup>		0.15	63	↑
121	fight	la <sup>55</sup> la <sup>33</sup>	intr.	0.04	22	↑
121	fight	tsu <sup>33</sup> tsu <sup>21</sup>	tr.	0.44	154	
122	five	wa <sup>33</sup>		0.19	88	↑
123	float	da <sup>21</sup>		0.48	161	
124	flow	ji <sup>33</sup>		0.46	159	
125	flower	ba <sup>33</sup> ba <sup>21</sup>		0.06	27	↑
126	fog	c <sup>h</sup> i <sup>55</sup> sɿ <sup>33</sup>		0.67	191	
127	four	lu <sup>33</sup>		0.17	77	↑
128	freeze	(ndzɿ <sup>21</sup> ) xa <sup>33</sup>		0.69	193	
130	grass	zi <sup>33</sup>		0.13	55	↑
131	guts	bv <sup>33</sup>		0.52	169	