

A Cross-linguistic Corpus of Forms Meaning ‘yes’

Steve Parker

SIL International and Graduate Institute of Applied Linguistics

Based on a carefully-compiled database of 604 attested forms for ‘yes’ taken from 512 languages spoken in over 70 countries, I show that this word exhibits a cross-linguistic tendency to contain laryngeal phonemes (/ʔ/ or /h/). As part of the statistical analysis I examine cognate items within specific genetic families and argue that certain phonotactic patterns involving ‘yes’ are not random in nature. These findings further corroborate the observation that glottal consonants often behave phonologically as a default or unmarked class of segments.

1. Introduction

A very basic and important aspect of natural human language is the fact that in the great majority of cases, the relationship between a word’s meaning and its pronunciation is arbitrary and unpredictable. Exceptions to this generalization of non-iconicity are therefore noteworthy. The purpose of this paper is to document a striking pattern I have discovered in languages from all areas of the world. Specifically, the lexical item meaning ‘yes’ has a fairly strong tendency to contain one or more glottal consonants — either [h] and/or [ʔ]. In the next section I present a corpus of forms listing the word(s) for ‘yes’ in 512 languages belonging to 64 major linguistic families and show that this phenomenon (laryngeal consonantism) is attested in at least 604 specific occurrences. In the ensuing discussion I give summary statistics and conclude that several common phonological themes occur with a frequency that is almost certainly greater than chance. The presentation here builds on previous work in Parker (1996, 2006). In the former paper I introduced the main pattern but was only able to include a truncated corpus of 44 forms due to limitations on space. And in the latter article (in Spanish) the organization of the word list by country obscures certain typological facts that are more directly elucidated here. The present paper constitutes the first full analysis in English of the entire corpus of 604 words, arranged and discussed according to genetic affiliation.

2. Data

In Table 1 below I list a series of lexical items meaning ‘yes’ in 512 specific languages. As noted above, the criterion for including a word in this corpus is any form for ‘yes’ which contains one or more instances of either or both glottal consonants — [h] and/or [ʔ], since this is the common pattern I have identified and propose to analyze here. I transcribe the items using IPA characters and generally repeat as much phonetic detail as each source reports. In a small number of cases it is not clear which (non-laryngeal) sound is being represented, so I simply reproduce the original symbols here, e.g., *ä*. Some of my sources transcribe the items phonetically, indicating complete surface realizations, while other sources use a more abstract, phonemic level of representation. However, since it is not always clear which of these two options is intended, I just faithfully copy each word below without indicating any distinction between different levels of phonological analysis. Nevertheless, there is one significant exception to this procedure which I consistently follow: in my data below I do not include any instances of word-initial [ʔ]’s when it is clear from the source that these are not phonemic. This is due to the very common cross-linguistic tendency for languages to epenthesize a phonetic [ʔ] as an automatic reflex to fulfill the

requirement for a syllable onset in phrase-initial or word-initial position. Consequently, since this nearly universal process would obviously confound my results here by greatly (and artificially) increasing the sample size, I exclude all such example words from my data. I include forms with a word-initial glottal stop only when it is clear from the source that that segment is contrastive in that position in that language. Therefore, all cases of initial [ʔ]'s in Table 1 below are assumed to be phonemic, as far as I am aware.

The data items I present here come from many different types of sources, and span over a decade of compilation. Long ago it became unwieldy to keep track of each reference, so I cannot list all of them in my bibliography section. Nevertheless, in all cases my preference is to rely on primary sources whenever possible. Consequently, the majority of these forms have been taken from published reference works such as dictionaries, descriptive grammars, etc. When feasible I also try to communicate directly in person with a linguist who has done extensive fieldwork on the language in question, or with a native speaker. However, for a relatively small number of cases I am not aware of a published source since at times I have included some data from places such as the Internet, survey reports by my SIL colleagues, etc. Consequently, it is not unlikely that a few transcriptional errors may have crept into my corpus. Nevertheless, given the overall robustness of the patterns I have observed in data from sources that are more reliable, none of my general conclusions are in doubt, as I will discuss in the next section.

Another issue which merits comment is the meaning of the items displayed in my list in Table 1 below. To the best of my knowledge, all of the words I present here are citation forms for 'yes' which are considered the standard, official way to express the concept of verbal assent, as in response to a yes/no question, for example. Many languages also have less formal equivalents, such as the English affirmation grunt typically written *uh-uh* (or in similar fashion). This type of expression is in fact very common, perhaps almost universal, so I have tried to filter it out of my corpus so as not to inappropriately inflate the statistics. Consequently, in the compilation of my word list I have purposely excluded all forms translated as 'yes' but which are specifically noted to be slang, informal, non-standard, etc. A related detail is that some languages do not have a single word exactly equivalent to 'yes,' but instead use a phrase meaning something like 'it is good.' In a very few cases of this type I have included such forms in my corpus, but always and only with the condition that the language in question must not have any other simpler and more direct way to express assent, and thus a published work such as a dictionary has listed this expression as the closest equivalent for 'yes.'

Before presenting the actual data, I should clarify that no attempt has been made to balance the sample of languages included here, either in terms of their linguistic affiliation or their areal locations, unlike the ideal list put together for typological purposes in *WALS* (Haspelmath et al. 2005; cf. Whaley 1997). Rather, Table 1 below includes every form for 'yes' I have discovered to date which meets the criterion spelled out above (a glottal consonant). As such, certain genetic families are represented very heavily, while certain others are not represented at all. Likewise, some continents have many languages with matching forms, while others have relatively few. This fact will make it difficult to extrapolate inferential statistics about the word 'yes' for the planet as a whole, but that is not my primary concern here. Rather, in offering this corpus I simply wish to document all the words for 'yes' with a laryngeal consonant that I am aware of, for the sake of exhaustivity. Consequently, there are hundreds of languages whose forms I have purposely excluded from this list, such as English *yes* and Spanish *sí*. In fact, the total number of languages I surveyed for this study was about 1372, of which 512 have one or more matching forms, so the overall hit rate for my sample is about 37%. After I give the attested forms I will return to these points and discuss them more systematically.

I now describe the internal structure of my corpus, as displayed in Table 1 below. For the spelling of language names and countries I follow the latest edition of the *Ethnologue* (Gordon 2005). I also follow this reference for the linguistic affiliation (genetic classification) of all the languages. (*Ethnologue* itself bases its organization of linguistic relationships on Frawley 2003.) The order of presentation of the languages in Table 1 is by family, following the geographical scheme of *WALS*, which in turn is derived from that of Ruhlen (1987). Within each first-order macro-group (phylum) or stock, the subfamilies are arranged alphabetically, again following *WALS*. Normally each family is broken down as far as the level of the genera posited by *WALS*, with a few minor deviations motivated by *Ethnologue*. After the name of each family, subfamily, and genus, I note in parentheses the number of languages from that group which appear in my corpus. Within each mini-table I list three pieces of information, from left to right: (1) the name of the language, (2) the official name of the country or countries where it is (or was) mainly spoken, and (3) the word or words meaning ‘yes,’ separated by commas. In cases when a language is spoken in more than one country, the one I list first is considered primary by *Ethnologue*. The order of the languages in the leftmost column of each mini-table is alphabetical.

Table 1: Corpus of forms meaning ‘yes’ (or ‘affirmation’)

Niger-Congo (20 languages), Atlantic-Congo (19 languages), Atlantic (1 language)

language	country	word(s) for ‘yes’
Jola-Fonyi	Senegal	ahej, ehe

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), Benue-Congo (10), Bantoid (9)

Akoose	Cameroon	ʔee, ʔěě
Digo	Kenya, Tanzania	ěhé
Fang	Equatorial Guinea	ěhě
Gikuyu	Kenya	eeh
Kwanyama	Angola, Namibia	heeno
Langi	Tanzania	ʔěhíé:
Mbala	Democratic Republic of the Congo	eʔe
Shona	Zimbabwe	ehe
Venda	South Africa, Zimbabwe	ih

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), Benue-Congo (10), Nupoid (1)

Nupe-Nupe-Tako	Nigeria	hin(jí)
----------------	---------	---------

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), Kru (1)

(Abu) Dida	Côte d’Ivoire	hěě
------------	---------------	-----

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), Kwa (3)

Akan	Ghana	ɛhɛɛ
Ga	Ghana	hɛ
Gen	Togo	heinn

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), North (4), Adamawa-Ubangi (2)

Mbum	Cameroon, Central African Republic	óʔó
Zande	Democratic Republic of the Congo, Central African Republic	hein

Niger-Congo (20), Atlantic-Congo (19), Volta-Congo (18), North (4), Gur (2)

Konni	Ghana	wǎʔ
Wali	Ghana	eʔe

Niger-Congo (20), Mande (1), Western (1)

Mandinka	Senegal	haa
----------	---------	-----

Afro-Asiatic (10), Berber (2)

Kabyle	Algeria	ih
Tachelhit	Morocco	ihe

Afro-Asiatic (10), Chadic (1), West (1)

Hausa	Nigeria	toh
-------	---------	-----

Afro-Asiatic (10), Cushitic (2)

Kambaata	Ethiopia	ʔǎǎ
Somali	Somalia	haa

Afro-Asiatic (10), Semitic (5)

Assyrian Neo-Aramaic	Iraq	he
Iraqi Arabic	Iraq	ʔii
Moroccan Spoken Arabic	Morocco	ih
Syrian (North Levantine Spoken) Arabic	Syria	ʔee
Tigrigna	Ethiopia	ʔuwej

Indo-European (23), Armenian (1)

(Eastern) Armenian	Armenia	ha
--------------------	---------	----

Indo-European (23), Celtic (1)

Scottish Gaelic	United Kingdom	haa
-----------------	----------------	-----

Indo-European (23), Indo-Iranian (19), Indo-Aryan (15)

Assamese	India, Bangladesh	haa
Bengali	Bangladesh	ha
Caribbean Hindustani	Suriname	han
Eastern Panjabi	India	ha(n) ji
Gujarati	India	haan
Hindi	India	hai, haʒa, ha(an)
Indus Kohistani	Pakistan	ah
Kashmiri	India, Pakistan	ho
Lambadi	India	hawə
Marathi	India	ho
Nepali	Nepal, India	haa
Panjabi	Pakistan, India	hāā
Romani	Romania	hai
Sindhi	Pakistan, India	ha
Urdu	Pakistan, India	hāā, ji hā, ha(ʒi)

Creole, Assamese based (Indo-European, Indo-Iranian, Indo-Aryan) (1)

Naga Pidgin	India	hoi
-------------	-------	-----

Indo-European (23), Indo-Iranian (19), Iranian (3)

Balochi	Pakistan, India	han
Central Kurdish	Iraq	hari
Pashto	Pakistan, Afghanistan	hoo

Indo-European (23), Slavic (2)

Slovak	Slovakia	hej
Upper Sorbian	Germany	haj

Uralic (1), Finnic (1)

Estonian	Estonia	jah
----------	---------	-----

Altaic (3), Turkic (3)

Azerbaijani	Azerbaijan	hæ
Turkmen	Turkmenistan	hawa
Uyghur	China	hāʔä

Japanese (1)

Japanese	Japan	hai(?), eei
----------	-------	-------------

North Caucasian (2)

Chechen	Russia	ha?
Ingush	Russia	hwa?a

Dravidian (1), Southern (1)

Kannada	India	haudu
---------	-------	-------

Sino-Tibetan (10), Chinese (2)

Hakka Chinese	China	hé
Yue Chinese	China	hai

Sino-Tibetan (10), Tibeto-Burman (8), Himalayish (Bodic) (4)

Chepang	Nepal	ma?
Limbu	Nepal	oo?
Newar(i)	Nepal	khah
Sherpa	Nepal	jeeah

Sino-Tibetan (10), Tibeto-Burman (8), Jingpho-Konyak-Bodo (1)

Chang Naga	India	háu
------------	-------	-----

Sino-Tibetan (10), Tibeto-Burman (8), Kuki-Chin-Naga (1)

Sumi Naga	India	ih
-----------	-------	----

Sino-Tibetan (10), Tibeto-Burman (8), Lolo-Burmese (2)

Akha	Myanmar, Thailand	ɲuh mah
Burmese	Myanmar, Bangladesh	hou?ke, hou?pade

Hmong-Mien (1)

Hmong	China, Thailand	huuv
-------	-----------------	------

Austro-Asiatic (10), Mon-Khmer (10), Aslian (5)

Jah Hut	Malaysia	jeh
Kensiu	Malaysia	hiʔih
(Perak) Semai	Malaysia	éh-é
(Ulu Kampar) Semai	Malaysia	hã
Temiar	Malaysia	tahatna

Austro-Asiatic (10), Mon-Khmer (10), Eastern Mon-Khmer (4), Bahnaric (3)

Bahnar	Viet Nam	höm öi, hám öi
Sedang	Viet Nam	hom
Stieng	Viet Nam	öh

Austro-Asiatic (10), Mon-Khmer (10), Eastern Mon-Khmer (4), Katuic (1)

Pacoh	Viet Nam	ʔu
-------	----------	----

Austro-Asiatic (10), Mon-Khmer (10), Northern Mon-Khmer (1), Khmuic (1)

Khmu	Laos, Viet Nam	he
------	----------------	----

Austronesian (133), Malayo-Polynesian (133), Bali-Sasak (1)

Sasak	Indonesia	aoʔ, auʔ
-------	-----------	----------

Austronesian (133), Malayo-Polynesian (133), Barito (Borneo) (3)

Dohoi	Indonesia	ijoʔ
Maʔanyan (Dayak)	Indonesia	hiʔai
Ngaju (Dayak)	Indonesia	joh

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Central Malayo-Polynesian (18), Aru (1)

Kola	Indonesia	ʔĩĩ
------	-----------	-----

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Central Malayo-Polynesian (18), Bima-Sumba (2)

Ende	Indonesia	oʔoh
Kambera	Indonesia	aʔa

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Central Malayo-Polynesian (18), Central Maluku (11)

Amahai	Indonesia	helo
Ambelau	Indonesia	ehe
Asilulu	Indonesia	ho-o
Boano	Indonesia	ode?
Buru	Indonesia	ehe
Elpaputih	Indonesia	iʔa
Geser-Gorom	Indonesia	helo
Saparua	Indonesia	ijawahi, hello
Sapolewa Seram	Indonesia	iʔjo, hɛʔɛ
Sepa	Indonesia	helo
Taliabu	Indonesia	ihi

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Central Malayo-Polynesian (18), Timor (4)

Bilba	Indonesia	hei
Sika	Indonesia	ehe
Tetun	Indonesia	hɛʔɛ, ho(u)
Uab Meto	Indonesia	hao, hé

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Admiralty Islands (15)

Bipi	Papua New Guinea	ɛhe
Kele	Papua New Guinea	heʔé, (e)'he
Khehek	Papua New Guinea	hɛʔɛ
Koro	Papua New Guinea	ehe
Kurti	Papua New Guinea	ehe
Leipon	Papua New Guinea	ɛhɛ
Lele	Papua New Guinea	ɛhɛʔ
Likum	Papua New Guinea	ehe
Loniu	Papua New Guinea	ɛhɛ
Lou	Papua New Guinea	saʔ
Mokerang	Papua New Guinea	'ɛhɛ
Mondropolon	Papua New Guinea	saʔ
Nali	Papua New Guinea	ɛʔhe
Nyindrou	Papua New Guinea	ɛhɛʔ
Wuvulu-Aua	Papua New Guinea	hiʔi

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), Remote Oceanic (13), Central Pacific (9), East Fijian-Polynesian (8)

Futuna-Aniwa	Vanuatu	ho
Hawaiian	United States	?ae
Maori	New Zealand	?aae, ?ee
Nukuria	Papua New Guinea	i'no?
Rarotongan	Cook Islands	?ae
Rennell-Belona	Solomon Islands	?oo
Samoa	Samoa	?oe, ?ii
Tongan	Tonga	?io

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), Remote Oceanic (13), Central Pacific (9), West Fijian-Rotuman (1)

Rotuman	Fiji	?i, ?o
---------	------	--------

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), Remote Oceanic (13), Micronesian (2)

Kosraean	Micronesia	ahok
Nauruan	Nauru	eh

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), Remote Oceanic (13), North and Central Vanuatu (2)

(East) Ambae	Vanuatu	ho?o
Sakao	Vanuatu	hao

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), South Vanuatu (3)

Aneityum	Vanuatu	ho
Kwamera	Vanuatu	owah
Lenakel	Vanuatu	ouaah

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Central-Eastern Oceanic (20), Southeast Solomonian (4)

Arosi	Solomon Islands	?a?a, ?e?e, ?uu
Bughotu	Solomon Islands	'hi?i, 'he?e
Kwaio	Solomon Islands	a?a
Kwara'ae	Solomon Islands	?iu

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Western Oceanic (33), Meso Melanesian (8), New Ireland (8)

Cheke Holo	Solomon Islands	heʔe
Halia	Papua New Guinea	geha
Kokota	Solomon Islands	ehe
Nehan	Papua New Guinea	'hawun
Petats	Papua New Guinea	oaiʔ
Saposa	Papua New Guinea	'ejεʔ
Solos	Papua New Guinea	ʔeh
Tinputz	Papua New Guinea	kèʔ

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Western Oceanic (33), North New Guinea (11), Huon Gulf (6)

Adzera	Papua New Guinea	hai
Bugawac	Papua New Guinea	aiʔ
Kela	Papua New Guinea	ʔεʔε
Wampar	Papua New Guinea	ʔijo
Yabem	Papua New Guinea	aeʔ
Zenag	Papua New Guinea	βaʔ

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Western Oceanic (33), North New Guinea (11), Ngero-Vitiaz (5)

Arop-Lokep	Papua New Guinea	εʔ
Bebeli	Papua New Guinea	eʔe
Gimi	Papua New Guinea	ehe
Karnai	Papua New Guinea	biɔʔ
Tami	Papua New Guinea	iʔ

Austronesian (133), Malayo-Polynesian (133), Central-Eastern (86), Eastern Malayo-Polynesian (68), Oceanic (68), Western Oceanic (33), Papuan Tip (14)

Anuki	Papua New Guinea	ʔeqa
'Auhelawa	Papua New Guinea	ehewa
Boselewa	Papua New Guinea	iʔwa
Buhutu	Papua New Guinea	ih
Bunamu	Papua New Guinea	'eh(e)wa
Doga	Papua New Guinea	ʔona
Duau	Papua New Guinea	éhe
Gumawana	Papua New Guinea	goʔ
Gweda	Papua New Guinea	hámada
Haigwai	Papua New Guinea	eʔeʔe
Iduna	Papua New Guinea	ehe
Keapara	Papua New Guinea	eʔe
Molima	Papua New Guinea	ʔao
Sewa Bay	Papua New Guinea	'eh

Austronesian (133), Malayo-Polynesian (133), Chamorro (1)

Chamorro	Guam, Northern Mariana Islands	huʔu
----------	--------------------------------	------

Austronesian (133), Malayo-Polynesian (133), Kayan-Murik (2)

Aoheng	Indonesia	haʔu
Busang Kayan	Indonesia	ioʔ

Austronesian (133), Malayo-Polynesian (133), Malayic (Sundic) (9)

Banjar	Indonesia, Malaysia	iʔih
Chru	Viet Nam	hèh
Jakun	Malaysia	jeh, iah, jaʔ
Jambi (Ulu) Malay	Indonesia	auʔ
Jarai	Viet Nam	hoi, hom
Pasemah	Indonesia	aʔu
Rade	Viet Nam	mah
Serawai	Indonesia	aʔu
Western Cham	Cambodia	hu, haij

Austronesian (133), Malayo-Polynesian (133), Meso Philippine (3)

Aklanon	Philippines	huo
Mansaka	Philippines	uʔu
Tagalog	Philippines	'o:ʔo

Austronesian (133), Malayo-Polynesian (133), Northwest (5), North Sarawakan (3)

Kelabit	Malaysia, Indonesia	heʔ-eh
Kenyah	Indonesia	ǎhàʔ
Tring	Malaysia	eʔa

Austronesian (133), Malayo-Polynesian (133), Northwest (5), Sabahan (2)

Dusun	Malaysia	oʔoh
Kadazan	Malaysia	oʔoh

Austronesian (133), Malayo-Polynesian (133), South Mindanao (1)

Tiruray	Philippines	hoʔo
---------	-------------	------

Austronesian (133), Malayo-Polynesian (133), Southern Philippine (1)

Dibabawon Manobo	Philippines	əʔə
------------------	-------------	-----

Austronesian (133), Malayo-Polynesian (133), Sulawesi (19)

Banggai	Indonesia	òʔò
Coastal Konjo	Indonesia	ioʔ
Dampelas	Indonesia	hije
Kulisusu	Indonesia	ũũhũ
Laiyolo	Indonesia	ijō-uh
Mori	Indonesia	huumbee
Padoe	Indonesia	humbe
(Petapa) Taje	Indonesia	hoʔo
Ratahan	Indonesia	u-hu
Selayar	Indonesia	ijō-uh
Suwawa	Indonesia	ooʔ
(Taruna) Sangir	Indonesia	eʔeŋ
Tolaki	Indonesia	oho
Tomini	Indonesia	ʔeie
Tondano	Indonesia	uhuʔ
Tontemboan	Indonesia	eʔen
Tukang Besi	Indonesia	oho
Waru	Indonesia	huŋ
Wawonii	Indonesia	hoo

Austronesian (133), Malayo-Polynesian (133), Sumatra (2)

Mentawai	Indonesia	oʔo
Nias	Indonesia	ahe, jaʔia

West Papuan (1), North Halmahera (1)

Galela	Indonesia	híaa
--------	-----------	------

Sko (2), Krisa (1)

Warapu	Papua New Guinea	'aʔo
--------	------------------	------

Sko (2), Vanimo (Western Sko) (1)

Skou	Indonesia	ʔæ
------	-----------	----

Torricelli (6), Kombio-Arapesh (3)

Bumbita Arapesh	Papua New Guinea	oʔuʔε
Wom	Papua New Guinea	auhe
Yambes	Papua New Guinea	oho

Torricelli (6), Marienberg (2)

Buna	Papua New Guinea	jooʔ
Kamasau	Papua New Guinea	eʔa

Torricelli (6), Wapei-Palei (1)

Urat	Papua New Guinea	he
------	------------------	----

Kwomtari-Baibai (1)

Baibai	Papua New Guinea	wəʔ
--------	------------------	-----

Left May (1)

Iteri	Papua New Guinea	wowoʔ
-------	------------------	-------

Sepik-Ramu (8), Ramu (2), Ramu Proper (2)

Arafundi	Papua New Guinea	ʔo
Kire	Papua New Guinea	aha

Sepik-Ramu (8), Sepik (6), Middle Sepik (2)

Kwoma	Papua New Guinea	hehe
Manambu	Papua New Guinea	haa-joú

Sepik-Ramu (8), Sepik (6), Sepik Hill (4)

Alamblak	Papua New Guinea	ʔoa
Bisis	Papua New Guinea	ʔɛʔej
Niksek	Papua New Guinea	i'pahe
Sumariup	Papua New Guinea	ʔejo

Trans-New Guinea (52), Main Section (32), Central and Western (23), Angan (1)

Baruya	Papua New Guinea	jaʔjo
--------	------------------	-------

Trans-New Guinea (52), Main Section (32), Central and Western (23), Central and South New Guinea-Kutubuan (3)

Bimin	Papua New Guinea	ʔa'o
Kasua	Papua New Guinea	'ɛhɛ
Konai	Papua New Guinea	hɛ'ɬæ

Trans-New Guinea (52), Main Section (32), Central and Western (23), East New Guinea Highlands (11), Central (1), Chimbu (1)

Kuman	Papua New Guinea	oʔo
-------	------------------	-----

Trans-New Guinea (52), Main Section (32), Central and Western (23), East New Guinea Highlands (11), East-Central (7)

Alekano	Papua New Guinea	ooʔ
Benabena	Papua New Guinea	óʔjo
Gende	Papua New Guinea	oʔo
Inoke-Yate	Papua New Guinea	he
Kanite	Papua New Guinea	he
Keyagana	Papua New Guinea	he
Yagaria	Papua New Guinea	he, hiβa

Trans-New Guinea (52), Main Section (32), Central and Western (23), East New Guinea Highlands (11), West-Central (3)

Angal	Papua New Guinea	ʔã
Angal Heneng	Papua New Guinea	ɛñ
Huli	Papua New Guinea	hee

Trans-New Guinea (52), Main Section (32), Central and Western (23), Huon-Finisterre (6)

Abaga	Papua New Guinea	oʔzo
Asaro'o	Papua New Guinea	goʔon
Awara	Papua New Guinea	hi'ʔi
Forak	Papua New Guinea	oʔ
Kâte	Papua New Guinea	ohoʔ
Mape	Papua New Guinea	o'oʔ

Trans-New Guinea (52), Main Section (32), Central and Western (23), Marind (2)

Kuni-Boazi	Papua New Guinea	eʔ
Zimakani	Papua New Guinea	aʔa

Trans-New Guinea (52), Main Section (32), Eastern (9), Central and Southeastern (9), Dagan (3)

Kanasi	Papua New Guinea	oʔa
Mapena	Papua New Guinea	ʔe
Turaka	Papua New Guinea	ʔe

Trans-New Guinea (52), Main Section (32), Eastern (9), Central and Southeastern (9), Goilalan (1)

Fuyug	Papua New Guinea	eʔe
-------	------------------	-----

Trans-New Guinea (52), Main Section (32), Eastern (9), Central and Southeastern (9), Koiarian (3)

Ese	Papua New Guinea	iʔa, kaʔivo
Grass Koiari	Papua New Guinea	nʔn, oʔe
Ömie	Papua New Guinea	iuʔu

Trans-New Guinea (52), Main Section (32), Eastern (9), Central and Southeastern (9), Kwalean (1)

Uare	Papua New Guinea	'oʔε
------	------------------	------

Trans-New Guinea (52), Main Section (32), Eastern (9), Central and Southeastern (9), Mailuan (1)

Mailu	Papua New Guinea	eʔe
-------	------------------	-----

Trans-New Guinea (52), Eleman (4)

Kaki Ae	Papua New Guinea	ɛhɛ
Opao	Papua New Guinea	ehe
Tairuma	Papua New Guinea	ahae
Toaripi	Papua New Guinea	aʔa

Trans-New Guinea (52), Madang-Adelbert Range (10), Adelbert Range (2)

Moresada	Papua New Guinea	oʔo
Tauya	Papua New Guinea	oʔo

Trans-New Guinea (52), Madang-Adelbert Range (10), Madang (8), Mabuso (5)

Garus	Papua New Guinea	ʔoʔ, æʔ
Girawa	Papua New Guinea	hoo
Rempi	Papua New Guinea	æʔ
Samosa	Papua New Guinea	oh
Wamas	Papua New Guinea	ʔuʔu

Trans-New Guinea (52), Madang-Adelbert Range (10), Madang (8), Rai Coast (3)

Ganglau	Papua New Guinea	oh
Sam	Papua New Guinea	oʔ
Yabong	Papua New Guinea	oʔo

Trans-New Guinea (52), Northern (3), Border (3)

Amanab	Papua New Guinea	ʔee
Sowanda	Papua New Guinea	jəəʔ
Waris	Papua New Guinea, Indonesia	ɔ̃ʔɔ̃

Trans-New Guinea (52), Trans-Fly-Bulaka River (3)

Bamu	Papua New Guinea	eʔe
Northeast Kiwai	Papua New Guinea	ʔεε
Waboda	Papua New Guinea	iʔo

East Papuan (3), Yele-Solomons-New Britain (1), New Britain (1), Kuot (1)

Kuot	Papua New Guinea	(ʔ)aa(ʔ)
------	------------------	----------

East Papuan (3), Bougainville (2), East (2)

Naasioi	Papua New Guinea	eeʔ
Sibe	Papua New Guinea	'euʔ

Australian (6), Pama-Nyungan (6)

Djinang	Australia	jaʔaw
Wik-Mungkan	Australia	eeʔ
Worimi	Australia (extinct)	njee-hu
Yugambal	Australia (extinct)	ŋeh

Australian (6), (Pama-Nyungan,) Kulin (2)

Colac (Gulidjan)	Australia	aha
Wathawurrung	Australia	aha, ha ha, eh eh

Eskimo-Aleut (1)

Pacific Gulf Yupik	United States	aaʔa
--------------------	---------------	------

Na-Dene (5), Nuclear Na-Dene (5), Athapaskan-Eyak (5)

Apache	United States	haʔoh, haʔah
Kato	United States (extinct)	heeʔuuʔ
Navajo	United States	aouʔ, aooʔ
Tanaina	United States	aaʔ
Tsetsaut	Canada (extinct)	haa ah

Algic (10), Algonquian (9)

Cheyenne	United States	héeheʔε, haáhe
Chippewa	United States	heh
Cree	Canada, United States	eʔheʔ, âha, îhî
Malecite-Passamaquoddy	Canada, United States	aha
Micmac	Canada, United States	ʔeehe, eʔe
Montagnais	Canada	ehe
Naskapi	Canada	niihiiij
Potawatomi	United States, Canada	eʔhe
Western Abnaki	Canada, United States	ôhô(ô)

Algic (10), Wiyot (1)

Wiyot	United States (extinct)	hè
-------	-------------------------	----

French-Cree mixed language (Indo-European, Italic, Romance + Algic, Algonquian) (1)

Michif	United States, Canada	aenhenk
--------	-----------------------	---------

Iroquoian (4), Northern Iroquoian (4)

Cayuga	Canada, United States	éhé
Mohawk	Canada, United States	hén
Seneca	United States, Canada	ʔεεʔ
Tuscarora	Canada, United States	heh-heh

Muskogean (3)

Alabama	United States	how
Choctaw	United States	āh
Muskogee	United States	henká, ho

Gulf (2)

Atakapa	United States (extinct)	ha(ha)
Chitimacha	United States (extinct)	aha

Siouan (7)

Biloxi	United States (extinct)	he
Catawba	United States (extinct)	himba
Dakota	United States	ha(n)
Hidatsa	United States	hao
Iowa-Oto	United States (extinct)	hunje
Lakota	United States	haw, han
Osage	United States	ho-

Kiowa Tanoan (2)

Jemez	United States	hah
Kiowa	United States	haaʔ

Uto-Aztecan (21), Northern Uto-Aztecan (10), Hopi (1)

Hopi	United States	asʔa, taʔa
------	---------------	------------

Uto-Aztecan (21), Northern Uto-Aztecan (10), Numic (6)

Comanche	United States	haa, hah
Kawaiisu	United States	huʔu
Mono	United States	haʔ, hühü
Northern Paiute	United States	aha, haʔa
Shoshoni	United States	hãã
Ute-Southern Paiute	United States	huʔú, hiʔi

Uto-Aztecan (21), Northern Uto-Aztecan (10), Takic (2)

Cahuilla	United States	hée
Luiseño	United States	ohoo

Uto-Aztecan (21), Northern Uto-Aztecan (10), Tubatulabal (1)

Tübatulabal	United States	han
-------------	---------------	-----

Uto-Aztecan (21), Southern Uto-Aztecan (11), Aztecan (2)

Pipil	El Salvador	eehe
Southeastern Puebla Nahuatl	Mexico	e'he

Uto-Aztecan (21), Southern Uto-Aztecan (11), Sonoran (9), Cahita (4)

Eudeve	Mexico	héve, heé, hoi éko
Mayo	Mexico	heewi
Opata	Mexico	haru
Yaqui	Mexico	héewi, hehe

Uto-Aztecan (21), Southern Uto-Aztecan (11), Sonoran (9), Corachol (2)

Cora	Mexico	hée
Huichol	Mexico	húu, húuu

Uto-Aztecan (21), Southern Uto-Aztecan (11), Sonoran (9), Tarahumaran (1)

Tarahumara	Mexico	húri
------------	--------	------

Uto-Aztecan (21), Southern Uto-Aztecan (11), Sonoran (9), Tepiman (2)

Pima Bajo	Mexico	heu?u
Tohono O'odham	United States, Mexico	huu?u, hau?u

Salishan (7), Central Salish (4)

Clallam	United States	?aa
Lushootseed	United States	?i
Southern Puget Sound Salish	United States	?i
Straits Salish	Canada, United States	hee?e

Salishan (7), Interior Salish (3)

Coeur d'Alene	United States	hej
Okanagan	Canada, United States	waj?
Spokane	United States	?a

Penutian (13), California Penutian (1), Wintuan (1)

Wintu	United States	ho(o), ʔume
-------	---------------	-------------

Penutian (13), Chinookan (1)

Chinook	United States	ah-ha e-eh
---------	---------------	------------

Penutian (13), Maiduan (1)

Maidu	United States	hee, heʔu
-------	---------------	-----------

Penutian (13), Plateau Penutian (2), Klamath-Modoc (1)

Klamath-Modoc	United States	ʔii
---------------	---------------	-----

Penutian (13), Plateau Penutian (2), Sahaptin (1)

Nez Perce	United States	ʔe-hé
-----------	---------------	-------

Penutian (13), Yok-Utian (8), Utian (7), Costanoan (1)

Ohlone	United States	he(ah)
--------	---------------	--------

Penutian (13), Yok-Utian (8), Utian (7), Miwokan (6)

Amador Miwok	United States	hu
Coast Miwok	United States	ʔúu
Mariposa Miwok	United States	huu
Plains Miwok	United States	hûû, he-la, hæʔə(h)
Southern Sierra Miwok	United States	huuuʔu
Tuolumne Miwok	United States	hu

Penutian (13), Yok-Utian (8), Yokuts (1)

Yokuts	United States	hò, hò(o)we, hò(u)hu, hûhu, hûn, hân, hòn(hu), houu
--------	---------------	---

Hokan (9), Esselen-Yuman (5)

Cocopa	Mexico, United States	ʔiiʔí, ʔãã
Esselen	United States (extinct)	iʔké
Havasupai-Walapai-Yavapai	United States	eʔ
Kiliwa	Mexico	ʔhaa
Kumiai	Mexico, United States	ʔe-en

Hokan (9), Northern (1), Karok-Shasta (1)

Achumawi	United States	há
----------	---------------	----

Hokan (9), Salinan-Seri (1)

Seri	Mexico	jo'ʔaa
------	--------	--------

Hokan (9), Tequistlatecan (1)

Chontal	Mexico	hé
---------	--------	----

Hokan (9), Washo (1)

Washo	United States	jeʔ
-------	---------------	-----

Yuki (2)

Wappo	United States (extinct)	ʔíʔih
Yuki	United States (extinct)	ʔããhãʔ, hãwhaʔ, ʔãh

Chumash (1)

Chumash	United States (extinct)	ho, hãʔme, ʔiʔ
---------	-------------------------	----------------

Oto-Manguean (13), Amuzgoan (1)

Amuzgo	Mexico	ʔaha
--------	--------	------

Oto-Manguean (13), Mixtecan (2)

San Miguel el Grande Mixtec(o)	Mexico	hãã
Santa María Zacatepec Mixtec(o)	Mexico	hùu

Oto-Manguean (13), Otopamean (4)

Atzingo Matlatzinca	Mexico	haa
Mazahua	Mexico	hã(gã)
Mezquital Otomi	Mexico	aha
Otomi	Mexico	hã(hã)

Oto-Manguean (13), Popolocan (3)

Ixcatec	Mexico	hã ² ã ³
Mazatec(o)	Mexico	hao
Popoloca	Mexico	haa

Oto-Manguenan (13), Zapotecan (3)

Mitla Zapotec(o)	Mexico	oʔ(n)
Tataltepec Chatino	Mexico	hwaʔã, tsoʔo
Zapotec(o)	Mexico	jaʔo

Totonacan (2)

Papantla Totonac(a/o)	Mexico	hé
Xicoteppec de Juárez Totonac(a/o)	Mexico	uʔwee

Mixe-Zoque (8)

Coatlán Mixe	Mexico	huuu
Copainalá Zoque	Mexico	huʔu
Francisco León Zoque	Mexico	huʔu
Mixe	Mexico	hadún
Oluta Popoluca	Mexico	hoo
Rayón Zoque	Mexico	huʔu
Sayula Popoluca	Mexico	hoo
Zoque	Mexico	ha(?)a

Huavean (1)

Huave	Mexico	aha(h)
-------	--------	--------

Mayan (18), Cholan-Tzeltalan (4)

Chol	Mexico	tʃeʔi
Ch'orti'	Guatemala	huhu
Tzeltal	Mexico	hitʃ
Tzotzil	Mexico	haʔ, hiʔ

Mayan (18), Huastecan (1)

Huastec(o)	Mexico	ohniʔ
------------	--------	-------

Mayan (18), Kanjobalan-Chujean (3)

Akateko (Western Q'anjob'al)	Guatemala	haaʔ
Eastern Q'anjob'al	Guatemala	haa
Tojolabal	Mexico	haʔi, oho

Mayan (18), Quichean-Mameam (7)

Ixil	Guatemala	he
K'iche'	Guatemala	heʔ
Mam	Guatemala	ho
Poqomchi'	Guatemala	ho
Q'eqchi'	Guatemala	eh he
Tacanec(o)	Guatemala, Mexico	oho-
Tektiteco	Guatemala	ʔoʔ, ʔu

Mayan (18), Yucatecan (3)

Itza'	Guatemala	haa
Lacandon	Mexico	laʔ
Mopán Maya	Belize, Guatemala	hah

Misumalpan (1)

Sumo-Mayangna	Nicaragua, Honduras	âwih
---------------	---------------------	------

Chibchan (2), Aruak (1)

Cogui	Colombia	aha
-------	----------	-----

Chibchan (2), Guaymi (1)

Ngäbere	Panama	hon
---------	--------	-----

Choco (2)

Epena	Colombia, Ecuador	óho
Woun Meu	Panama, Colombia	ʔeera

Barbacoan (1), Cayapa-Colorado (1)

Chachi	Ecuador	heen
--------	---------	------

Guahiban (1)

Guahibo	Colombia, Venezuela	hãhãʔ
---------	---------------------	-------

Tucanoan (8)

Carapana	Colombia, Brazil	ãhã, hau
Cubeo	Colombia, Brazil	hu
Desano	Brazil, Colombia	ãʔã
Koreguaje	Colombia	hĩhĩ
Secoya	Ecuador, Peru	hau, huĩhuʔu
Tanimuca-Retuarã	Colombia	ãʔã
Tatuyo	Colombia	ʰhʌu(?)
Tucano	Brazil, Colombia	hai

Witotoan (3), Boran (1)

Bora	Peru	hééé, huúúhu
------	------	--------------

Witotoan (3), Witoto (2)

Murui Huitoto	Peru	hi, huuu, hee
Ocaina	Peru	híí, huuu, hãã

Zaparoan (1)

Arabela	Peru	hãã
---------	------	-----

Peba-Yaguan (1)

Yagua	Peru	hoo
-------	------	-----

Jivaroan (2)

Achuar-Shiwiar	Peru	haʔaj
Aguaruna	Peru	uʔũ

Cahuapanan (1)

Chayahuita	Peru	iʔi
------------	------	-----

Panoan (7)

Amahuaca	Peru	huũʔũ
Capanahua	Peru	huúú, hóó
Cashinahua	Peru, Brazil	haa, huĩ
Panobo	Peru	huĩhuĩ
Shipibo-Conibo	Peru	huĩhuĩ
Yaminahua	Peru	ũhuũ
Yora	Peru	uuhũ

Quechuan (2)

Arequipa-La Unión Quechua	Peru	õʔ
Inga	Colombia	aha

Aymaran (2)

Aymara	Peru	his(a)
Jaqaru	Peru	haa

Harakmbet (1)

Amarakaeri	Peru	ẽẽʔ
------------	------	-----

Maku (2)

Hupdë	Brazil, Colombia	hʌʔ
Yuhup	Brazil	hʌʔ

Arawakan (18), Maipuran (18)

Asháninka	Peru	he
Ashéninka	Peru	hěě
Ashéninka Pajonal	Peru	hěě
Baure	Bolivia	hah
Caquinte	Peru	'hěěhě
Chamicuro	Peru	'ěhě
Ignaciano	Bolivia	heʔe, (ha)ʔá
Iñapari	Peru	ahamá
Machiguenga	Peru	'hěěhe, ne'ʔee
Nomatsiguenga	Peru	heé
Parecís	Brazil	hahan
Resígaro	Peru	háke
Taino	Bahamas (extinct)	han(-haʔn)
Tariano	Brazil	háw
Wayuu	Colombia, Venezuela	ah(á)
Yanesha'	Peru	hãã
Yine	Peru	hũhũ
Yucuna	Colombia	áʔa

Carib (1)

Wayana	Suriname	ihi, ehë
--------	----------	----------

Tupi (11), Arikem (1)

Karitiâna	Brazil	hãã
-----------	--------	-----

Tupi (11), Mawe-Satere (1)

Sateré-Mawé	Brazil	'taaʔi
-------------	--------	--------

Tupi (11), Tupi-Guarani (9)

Avá-Canoeiro	Brazil	hiba
Guajajára	Brazil	hê-, aʔê
Guaraní	Brazil, Bolivia, Argentina	hõo, hãa, haʔe, hee
Kamayurá	Brazil	heʔen
Tembé	Brazil	hẽ'ʔẽ
Tenharim	Brazil	haʔã
Urubú-Kaapor	Brazil	hã, aʔé
Wayampi	French Guiana, Brazil	õʔõ
Zo'é	Brazil	ehe

Macro-Ge (5), Ge-Kaingang (4)

Kaingáng	Brazil	hã
Xavánte	Brazil	ĩhe
Xerénte	Brazil	ĩhe, 'ehe
Xokleng	Brazil	hõ

Macro-Ge (5), Maxakali (1)

Maxakalí	Brazil	'hãʔõ
----------	--------	-------

Nambiquaran (1)

Nambikuára	Brazil	hàjó
------------	--------	------

Arauan (3)

Culina	Brazil, Peru	heʔe
Paumarí	Brazil	haʔa
Suruahá	Brazil	hiza

Tacanan (4)

Araona	Bolivia	hehe
Cavineña	Bolivia	heheʔe
Ese Eija	Bolivia, Peru	eʔe
Tacana	Bolivia	hadé, haʔá, (h)eʔe

Mataco-Guaicuru (2)

Abipon	Argentina (extinct)	haa, hee
Chorote	Argentina, Bolivia	xaʔe

Isolates (5)

Candoshi-Shapra	Peru	(m)a'ʔaa
Itonama	Bolivia	ãha
Kutenai	Canada, United States	hê
Urarina	Peru	ẽhẽ
Zuni	United States	haugh

3. Analysis and discussion

The table just presented lists a total of 604 words for ‘yes’ taken from 512 languages belonging to 64 major linguistic families, including five isolates. In this section I give summary statistics and highlight several interesting phonological patterns evident in the data. As noted in §2, no attempt was made to balance this sample either genetically or geographically; rather, it is a complete list of every matching form I have discovered to date. Hence, certain families are represented quite adequately, such as Austronesian with 133 languages, while others are notoriously absent. For example, there is not a single language from the Nilo-Saharan stock in my corpus. (In this paper I use the terms *phylum* and *stock* interchangeably.) This outcome is not due to any intentional purpose on my part; rather, it is a more or less accidental consequence of which parts of the world I have worked in and the concomitant collection of libraries I have had access to. In the compilation of my corpus I never avoided researching certain families or areas just because I suspected they would produce meager results. So while the sample of languages I explored is not completely random, neither is it biased in any obvious and predetermined way that would invalidate the results here.

Having clarified this point, I also now note that the relative distribution of languages in my corpus is in fact fairly well spread out among the major stocks and areas of the world. I document this in Table 2 below. From left to right I list the name of the major linguistic family, then the number of languages in that group which appear in my sample, followed by the total number of languages in that family according to *Ethnologue*, and finally, the corresponding percentage (number of languages from that phylum in my sample compared with total number of member languages in *Ethnologue*). In this table I only mention major families represented by ten or more languages in my data, and arrange them numerically from highest to lowest:

name of major stock	number of languages in my corpus	total number of member languages (<i>Ethnologue</i>)	percentage
Austronesian	133	1246	10.7%
Trans-New Guinea	52	561	9.3
Indo-European	23	430	5.3
Uto-Aztecan	21	56	37.5
Niger-Congo	20	1495	1.3
Mayan	18	68	26.5
Arawakan	18	49	36.7
Penutian	13	23	56.5
Oto-Manguean	13	172	7.6
Tupi	11	60	18.3
Afro-Asiatic	10	353	2.8
Sino-Tibetan	10	399	2.5
Austro-Asiatic	10	169	5.9
Algic	10	31	32.3
(overall totals)	362	5112	7.1%

Table 2: Linguistic families containing at least 10 languages in my database (taken from Table 1)

In analyzing Table 2 above, it should be emphasized that the figures in column three (total number of member languages) represent the hypothetically largest possible sample sizes for those families in the world, assuming that we had available to us the corresponding data (the words for 'yes') from each language. In actual practice I was not able to exhaustively survey any of these families, so the percentages in column four correspond to preliminary hit rates (proportion of languages with a matching form) for my corpus, at an absolute minimum, i.e., assuming the complete sample sizes in column three. I am not able to supply the real hit rates per family for my study, unfortunately, since I did not keep close track of the genetic affiliations of the languages I surveyed which did *not* exhibit matching words for 'yes' (forms with a glottal consonant). All that I tabulated was the approximate number of misses, which added up to about 860 languages. Consequently, the complete sample size for the planet as a whole (in this paper) is roughly 1372 languages surveyed, of which the total number displayed in Table 1 (512) equals an overall matching rate of about 37.3%. The quantity of languages for which I was able to ascertain the word for 'yes' (1372) corresponds to a 19.8% sample of all the living languages in the world (6912), according to *Ethnologue*. This is a fairly robust figure given the magnitude of the task.

Returning now to Table 2, if my data on all the languages in the world were exhaustive, the final percentages (hit rates) in column four would all potentially increase, although to what degree is hard to know for sure. As it stands, the highest actually attested proportion (among families with ten representatives or more) is 56.5% for the Penutian stock (13 matching languages out of 23 extant). This is encouraging. On the other hand, the family with the lowest hit rate in Table 2 is Niger-Congo (1.3%). This is symptomatic of the relatively low level of access I have had to data on African languages in general (so far). At the same time, it is not surprising that the two most numerous families in my corpus — Austronesian and Trans-New Guinea with 185 combined languages — are located in the part of the world where there is greatest linguistic diversity and density (the South Pacific). The overall number of first-order

families exemplified by at least one language in my corpus is 64, which amounts to 68.1% of the 94 total posited by *Ethnologue*. This too is a promising indicator.

I now move on to discuss a few aspects of the phonological content of the 604 words in my corpus in Table 1. The total number of glottal consonants in all forms combined is 761, so on average each word contains about 1.3 laryngeals. Of these, 474 or 62.3% consist of [h], while the remaining 287 (37.7%) are [ʔ]. The ratio of [h] to [ʔ] then is roughly 3:2. Among all these occurrences, [h] appears word-initially in 290 forms (61.2%); the remaining 184 tokens of [h] (38.8%) are non-initial. So [h] prefers initial over non-initial position by a margin of almost 2-to-1. Indeed, nearly one-half of all the words for ‘yes’ in my database begin with [h]. As far as [ʔ] is concerned, only 64 of its tokens are word-initial (22.3%), while the remaining 223 occurrences (77.7%) are non-initial. So [ʔ] prefers non-initial position over initial by a margin of almost 4-to-1. This is probably related to the fact that phonemic /ʔ/’s in general tend not to occur word-initially in many languages anyway.

At this point we might entertain the question, with what degree of statistical confidence can we now posit that these tendencies are significantly greater than chance? Although this issue is an important one, I am not in a position to answer it conclusively here, for two main reasons: (1) the list of data in Table 1 does not equally cover all linguistic families and geographic locations, and (2) even if my sample were ideally balanced, any global inferential test would be undermined by the fact that we don’t know the actual hit-or-miss rates for each phylum of languages. In retrospect this was an unfortunate methodological oversight on my part. In a perfect world, where we had exhaustive data on every language and could thus calculate the proportion of matching forms for any subset of languages, we would be able to proceed by comparing cognate words for ‘yes’ within each lowest-level genetic grouping, reconstruct the corresponding proto-form and its rate of retention in each daughter language, and then work our way backwards and up each higher-order branch of the tree until we could make a definitive generalization about each stock of related languages. Obviously this is not possible in the present case, so absolute statistical probabilities, as in works such as Ringe (1995), will have to wait for future research. As it stands, the chances of getting x number of look-alike hits in a large sample like this increases greatly when the corpus contains many related languages, as mine does. On the other hand, since many of the non-matching languages that I surveyed were also related to each other, this would tend to pull down the hit rates. Nevertheless, we cannot assume that these two opposing factors cancel each other out in any meaningful way, even if we could calculate them exactly. So the percentage figures I give above for the relative frequencies of [h] and [ʔ] should only be considered very rough estimates of the corresponding population rates (for all the languages in the world). This is especially true since an expression that sounds like *uh-uh*, for a concept that means something like ‘yes,’ is highly susceptible to being borrowed from neighboring ethnic groups by diffusion, even if the languages are not related. What is more, in any cross-linguistic comparison of this type, a certain percentage of apparent cognates will always occur by chance no matter what (Ringe 1995). Nevertheless, having noted these caveats, we can still at the very least make a few tentative predictions or claims about what we should reasonably expect to find among the remaining languages of the world:

- (1) Hypothesis 1: All else being equal, if the word for ‘yes’ in a particular language contains a laryngeal consonant, this is more likely to be [h] than [ʔ].

Hypothesis 2: All else being equal, if the word for 'yes' in a particular language contains an [h], this is more likely to be word-initial than non-initial.

Hypothesis 3: All else being equal, if the word for 'yes' in a particular language contains a [ʔ], this is more likely to be non-initial than initial.

At this point I note that the three predictions in (1) above may not necessarily be specific to the word for 'yes,' but rather may derive from more general patterns among the lexicons of the world's languages. For instance, the tendency of [ʔ] to avoid word-initial position across the board was already mentioned (cf. hypothesis 3). With respect to the preference for [h] to occur morpheme-initially (cf. hypothesis 2), this is actually enforced as a grammatical constraint on the occurrence of [h] in most lexical items in many languages: English (Davis 1999), Cuzco Quechua (Parker and Weber 1996), Panobo or Huariapano (Parker 1994), etc. Finally, let us consider hypothesis 1, whereby [h] is preferred over [ʔ] by a proportion of about 3:2 in this sample. This fact may simply be a reflection of the universal tendency of /h/ to appear more often than /ʔ/ does in phonemic inventories cross-linguistically. For example, in the UPSID database of 451 languages (Maddieson and Precoda 1992), /h/ occurs 279 times (61.9%) and /ʔ/ 216 times (47.9%). Similarly, in the P-base sample of 549 languages (Mielke 2006), /h/ appears in 361 inventories (65.8%) and /ʔ/ in only 195 (35.5%). While these latter two samples are not as ideally balanced as *WALS* is, their convergence nevertheless allows us to reasonably posit that /h/ is probably more frequent as a phoneme in the world's languages than /ʔ/ is. In a sense, then, the three hypotheses in (1) are completely natural and expected.

In order to go a step further and precisely quantify these three tendencies (from (1) above), technically speaking we would really need to know the phonemic inventory of every language studied, as well as the relative frequencies of each segment in each language-specific lexicon. This monumental task is beyond the scope of this study, and is not necessary for our purposes here. Nevertheless, keeping in mind the disclaimers above about the unbalanced nature of my sample, we still have enough data to arrive at some concrete conclusions for a few of the major families from Table 2. For each stock represented by ten or more languages in my database, I counted up the total number of [h]'s and [ʔ]'s among all their matching forms, ignoring the position of these sounds in the words where they occur. I then calculated (by phylum) the probability that the preference for one segment or the other is significantly greater than chance, using the binomial cumulative distribution (two-tailed). A similar result could also be obtained with a chi-squared test. Both of these procedures tend to be unreliable with samples consisting of less than ten tokens. In Table 3 below I display the results for those families which yielded significant results. To control for the effect of multiple comparisons (type 1 errors), I use a Bonferroni adjustment and test each contrast at an α level of .0036, which was arrived at by dividing .05 by 14 (the number of families listed in Table 2). Given this criterion, only five genetic groups have a preference for [h] or [ʔ] extreme enough — and with enough tokens — to be reliable. In the following table I arrange these families by p value, from lowest to highest:

family	<i>h</i>	<i>ʔ</i>	<i>p</i>
Indo-European	27	0	.0000
Penutian	28	7	.0005
Arawakan	24	5	.0005
Uto-Aztecan	32	10	.0009
Trans-New Guinea	17	43	.0011

Table 3: Language families in Table 2 which have a significant preference for one glottal consonant over the other one

As indicated in Table 3, the Indo-European languages overwhelmingly prefer to express their word for ‘yes’ with [h]. Every single Indo-European example in my sample contains exactly one [h] and no [ʔ]’s. Undoubtedly this is related to the fact that few languages in this family have the phoneme /ʔ/ at all. The only major stock which has a significant overall preference for [ʔ] over [h] is Trans-New Guinea. In addition to these generalizations, there are a few other trends we can note for some of the smaller families, even though they are not statistically significant. The three Altaic words all begin with [h] and the three East Papuan words end with [ʔ]. All eight Siouan words begin with [h] and lack [ʔ]’s completely. The four Yuki words each contain both laryngeal consonants. The eight Mixe-Zoque forms all begin with [h], as do the eight Witotoan words. Every Panoan language has a form containing the syllable [hu]. Every Macro-Ge and Arauan word contains an [h].

In addition to the tendency for the word meaning ‘yes’ to contain one or more glottal consonants, there is another indication that these forms are somewhat special cross-linguistically in another way as well: in many cases the [h] or [ʔ] is exceptional in that its occurrence is prohibited in the language as a whole, or at least highly restricted. I document some of these anomalies below (following the order of Table 1):

language	family	‘yes’	constraint
(East) Ambae	Austronesian	hoʔo	only word with [ʔ]
Lenakel	Austronesian	ouaah	only word with final [h]
Arop-Lokep	Austronesian	εʔ	only three other words with [ʔ]
Skou	Sko	ʔæ	only word with [ʔ]
Awara	Trans-New Guinea	hiʔi	only word with an intervocalic [ʔ]
Grass Koiari	Trans-New Guinea	nʔn, oʔe	only words with [ʔ]
Kuot	East Papuan	(ʔ)aa(ʔ)	only word with [ʔ]
Djinang	Australian	jaʔaw	only word with [ʔ]
Micmac	Algic	ʔeehe	only two other words with [h]
Montagnais	Algic	ehe	only three other words with [h]
Achuar-Shiwiar	Jivaroan	haʔaj	only word with an intervocalic [ʔ]
Panobo	Panoan	hũhũ	only word with an intervocalic [h]
Chamicuro	Arawakan	ʔẽẽ	only word with an intervocalic [h]
Yanashaʔ	Arawakan	hãã	only word with [h]
Candoshi-Shapra	Isolate	(m)aʔaa	only word with an intervocalic [ʔ]

Table 4: Languages having special restrictions on laryngeal consonants in general

Another case analogous to the examples in Table 4 above is provided by the English expression *uh-uh*. This is one of the few forms in the language in which the phoneme /h/ occurs in the middle of a morpheme; usually /h/ is restricted to morpheme-initial position. One other unusual detail about this word, for English, is that it is normally pronounced with nasalized vowels, even though these are not adjacent to a true nasal consonant like /m/ or /n/. This is a

classic illustration of the phenomenon of rhinoglottophilia, which Matisoff (1975:265) defines as “an affinity between the feature of nasality and the articulatory involvement of the glottis” (cf. Parker 1996, 2006). (In general this seems to be more frequent with /h/ than with /ʔ/.) This type of irregular nasalization is also common in my database in Table 1, where 64 words (10.6% of the total) have at least one nasalized vowel. What I do not know is whether this amount is significantly higher than the rate of occurrence of nasalized vowels overall in these languages, or for that matter in the whole world (in words other than ‘yes’). Nevertheless, several of my sources for this study point out that the word for ‘yes’ in particular languages exceptionally contains the only contrastive or unpredictably nasalized vowel(s) in the entire lexicon. In the following table I list those cases which I have noted to date:

language	family	‘yes’
Kambaata	Afro-Asiatic	ʔãã
Azerbaijani	Altaic	hã
Kola	Austronesian	ʔĩĩ
Shoshoni	Uto-Aztecan	hãã
Ashéninka	Arawakan	hẽẽ
Ashéninka Pajonal	Arawakan	hẽẽ
Chamicuro	Arawakan	ʔẽẽ
Yanesha’	Arawakan	hãã

Table 5: Languages in which nasalized vowels are restricted to the word for ‘yes’

Before closing this discussion I have a few comments to make about vowel quality in general (not just oral vs. nasal). While this paper has focused primarily on consonants, there are also several vowel patterns which form nice generalizations. For the five universally unmarked cardinal vowels, I counted up the number of words in my corpus in which each one is the first nuclear segment. I present the results in the table below, in which I also indicate the corresponding percentage of the total of 604 words:

segment	number of forms as first vocalic mora	percentage of total words
a	188	31.1%
e	149	24.7
o	96	15.9
i	63	10.4
u	29	4.8
totals	525	86.9

Table 6: Relative frequencies of the five cardinal vowels in the corpus in Table 1

As Table 6 shows, unrounded vowels tend to be more preferred than rounded ones, which is phonologically natural — lip rounding entails an additional articulatory gesture (de Lacy 2002). Also, within each of these two sets, lower (more sonorous) vowels are more frequent than higher ones. These two tendencies joined together converge on a significant (non-random) preference for the vowel /a/ in the word for ‘yes’ cross-linguistically ($\chi^2_{(4)} = 156.6$, $p < .0000$). This is hardly surprising since /a/ is universally unmarked anyway (de Lacy 2002, 2004). Furthermore, pharyngeal and glottal consonants tend to induce lowering on adjacent vowels in general, a well-known type of allophonic or morphophonemic conditioning via spreading (Kenstowicz 1994, McCarthy 1994).

The last item of business is simply to list some of the most common forms in my corpus. The following table displays the eight most frequent variants of the word for ‘yes’ in my data, ignoring minor (secondary) details such as vowel nasalization, stress, and tone. They are ordered by decreasing number of occurrences in my database, and are exhaustive in the sense that I have not tried to balance this table by limiting the tokens to only one exemplar per family:

form	number of occurrences
ehe	26
haa	25
he	20
ha	15
aha	13
hee	10
eʔe	10
aʔa	7

Table 7: Relative frequencies of the most common patterns for the word ‘yes’ in Table 1

The canonical forms in the table above nicely summarize and illustrate the general themes I have described throughout this section.

4. Conclusion

In any scientific endeavor, the most important question we can ask ourselves is, why should the world be the way it is? In this case, why should there be a universal tendency for the word meaning ‘yes’ to contain one or more glottal consonants? One factor which undoubtedly helps to explain this phenomenon is the fact that the laryngeal place of articulation node is inherently unmarked (Lombardi 2001, 2002), based on its typical phonological behavior as placeless (Halle 1995, Ladefoged 1997, Parker 2001). In summary, *Yes!* there is something interesting going on here cross-linguistically, and it clearly appears to exceed random chance. That is, we have probably discovered a worldwide articulatory pattern that maps meaning onto sound in a non-arbitrary way in many languages.

Acknowledgements

This paper has received very helpful input and suggestions from many people in many places at many times. In particular, though, I would like to thank two anonymous reviewers, as well as audiences at the University of Oregon, the University of Technology in Lae (Papua New Guinea), the University of North Dakota, and the Universidad Ricardo Palma in Lima, Peru.

References

- Davis, Stuart. 1999. The parallel distribution of aspirated stops and /h/ in American English. Indiana University working papers in linguistics 1.1-10.
- de Lacy, Paul. 2002. The formal expression of markedness. Ph.D. dissertation. University of Massachusetts Amherst.

- de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21/2.145-99.
- Frawley, William J. (ed.) 2003. *International encyclopedia of linguistics* (second edition). Oxford: Oxford University Press.
- Gordon, Raymond G., Jr. (ed.) 2005. *Ethnologue: languages of the world* (fifteenth edition). Dallas: SIL International.
- Halle, Morris. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26/1.1-46.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), with the collaboration of Hans-Jörg Bibiko, Hagen Jung, and Claudia Schmidt. 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Kenstowicz, Michael. 1994. *Phonology in generative grammar*. (Blackwell Textbooks in Linguistics.) Cambridge, Massachusetts and Oxford, UK: Blackwell.
- Ladefoged, Peter. 1997. Linguistic phonetic descriptions. *The handbook of phonetic sciences*, ed. by William J. Hardcastle, and John Laver. Oxford, UK and Cambridge, Massachusetts: Blackwell. pp. 589-618.
- Lombardi, Linda. 2001. Why place and voice are different: constraint-specific alternations in optimality theory. *Segmental phonology in optimality theory: constraints and representations*, ed. by Linda Lombardi. Cambridge: Cambridge University Press. pp. 13-45.
- Lombardi, Linda. 2002. Coronal epenthesis and markedness. *Phonology* 19/2.219-51.
- Maddieson, Ian, and Kristin Precoda. 1992. UPSID. Los Angeles: UCLA phonetics laboratory.
- Matisoff, James A. 1975. Rhinoglottophilia: the mysterious connection between nasality and glottality. *Nasálfest* (papers from a symposium on nasals and nasalization), ed. by Charles A. Ferguson, Larry M. Hyman, and John J. Ohala. Stanford: Language Universals Project, Department of Linguistics, Stanford University. pp. 265-87.
- McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. *Phonological structure and phonetic form: papers in laboratory phonology III*, ed. by Patricia A. Keating. Cambridge: Cambridge University Press. pp. 191-233.
- Mielke, Jeff. 2006. P-base. <http://www.u.arizona.edu/~mielke/research/pbase.html>
- Parker, Steve. 1994. Coda epenthesis in Huariapano. *International Journal of American Linguistics* 60/2.95-119.
- Parker, Steve. 1996. Toward a universal form for 'yes': or, rhinoglottophilia and the affirmation grunt. *Journal of Linguistic Anthropology* 6/1.85-95.
- Parker, Steve. 2001. Non-optimal onsets in Chamicuro: an inventory maximised in coda position. *Phonology* 18/3.361-86.
- Parker, Steve. 2006. La rinoglotofilia y el gruñido de afirmación — una tendencia universal. *Lengua y sociedad* 8/1.27-56.
- Parker, Steve, and David Weber. 1996. Glottalized and aspirated stops in Cuzco Quechua. *International Journal of American Linguistics* 62/1.70-85.
- Ringe, Donald A., Jr. 1995. 'Nostratic' and the factor of chance. *Diachronica* 12/1.55-74.
- Runner, Jennifer. 2003. "Yes" in over 550 languages. <http://www.elite.net/~runner/jennifers/yes.htm>
- Ruhlen, Merritt. 1987. *A guide to the world's languages, volume 1: classification*. Stanford: Stanford University Press.
- Whaley, Lindsay J. 1997. *Introduction to typology: the unity and diversity of language*. Thousand Oaks, California: Sage Publications.