

Power-laws and preferential attachment in linguistic morphology: Evidence from Meʔpá verb classes

Søren WICHMANN

Leiden University & Kazan Federal University & Beijing Language University

Abstract: Through extensive analysis it can be shown that the organization of the highly complex Meʔpá (Azoyú Tlapanec) verbal paradigms is for the most part transparently based on the expression of grammatical categories, even if these categories are sometimes typologically unfamiliar. There are also phonological principles at work in defining various subclasses, even if these are sometimes also difficult to unravel. However, within one of the grammatically defined classes it is necessary to operate with subclasses that are truly arbitrary. This concerns the tonal paradigms of semantically transitive verbs taking a high-impact (ergative) animate agent and an inanimate undergoer. For these, person inflection is tonal, and 24 different tonal paradigms have been identified. Apart from negative evidence –absence of non-arbitrary phonological, semantic or grammatical determinants of class membership– it is also possible to observe positive evidence that class membership is essentially stochastic. First, the sizes of the classes follow a power-law distribution (with a few large classes, some intermediate ones and many small ones, resulting in a straight line having a negative slope when size is plotted as a function of rank of size using log-log scales). Secondly, the more members a tonal class has, the more similar it is to other classes. Together, these two characteristics bear witness to a diachronic trajectory involving the principle of preferential attachment, which also governs many natural and social phenomena.

Keywords: Tlapanec, Meʔpá, Oto-Manguan, power laws, verbal inflection, tonal classes, ergative

1. Introduction

Arbitrary membership in verb classes can be defined through a battery of negative diagnostics:

- membership is not predictable from phonological characteristics;
- membership is not predictable from semantics;
- membership is not tied to a particular grammatical function.

If these conditions hold it would seem that one would be confronted with totally arbitrary class membership with no predictions to be made about the

behavior of the various classes. I will suggest, however, that the story does not end here. There are two additional positive characteristics which, if the set of classes is large enough, are expected to emerge:

- the number of members in a class is proportional to the formal similarity of the class to other classes; in other words, the larger the class, the more it resembles other classes on average;
- the sizes of classes exhibit a power-law distribution.

These two characteristics, which I interpret as characteristic of a stochastically driven linguistic subsystem, were found by observing one particular group of Meʔpá verbs. As far as I know they have not previously been detected in morphological studies, but I will be arguing that they are unsurprising, so I would also expect them to be widespread and to be good diagnostics of pure inflectional classes, i.e., classes that are not defined phonologically, semantically or grammatically.

After a section where I introduce the aspects of Meʔpá grammar that are most relevant for contextualizing the paradigms that we will be interested in, I go on to present the data on verb classes. A brief discussion of the implications of the findings follows.

2. Relevant aspects of Meʔpá grammar

In order to refer to the language presently under consideration I will use the auto-denomination Meʔpá.¹ The language makes use of seven tones, symbolized here by diacritical marks (á = H; a = M; à = L; ǎ = MH; ǎ = LM; â = HM; â = ML). Tones have a heavy functional load, especially in the grammatical domain. The language is head-marking, predicates being marked for person, number, agentivity (Wichmann 1996), one of four different case relations (ergative, absolutive, ‘pegative’, and dative, cf. Wichmann 2005a, 2009), one of three different aspects (perfective, imperfective, potential), polarity (affirmative, negative), and, finally, a

¹ In earlier publications I have used Azoyú Tlapanec, where Azoyú refers to the specific place where the language is spoken, near the coast of the Mexican state of Guerrero, and Tlapanec designates a group of closely related languages forming a subgroup of the Oto-Manguean family. The auto-denomination, however, is becoming more frequent in the literature, and also has the advantage that its phonological shape seems to vary across languages and dialects in such a way as to be indicative of which variant one is dealing with. For instance, the name of the Malinaltepec variant recorded by Suárez (1983) is Miʔphà:. Most of the data used in this paper was collected during 1991-1994 as part of my PhD research but is published here for the first time.

reference-tracking mechanism distinguishing between new and given third person animate participants (Wichmann 2007). The language does not make any overt distinctions pertaining to transitivity: morphologically there is no distinction between a transitive and an intransitive verb and there are no productive derivations involving transitivity, such as passive, antipassive or causative. Valency, however, is important, and what matters here is the number of animate participants, which can range from zero to three. Verbs do not undergo derivations serving to increase or decrease their valency, instead they are simply plugged into the inflectional paradigm appropriate for a given valency. For instance, *na-ndrèhmé* means ‘s/he is sprinkling it’ and *na-ndrihmá* means ‘s/he is sprinkling it on someone’. The former pertains to a paradigm of monopersonal verbs encoding the ergative case role and the latter to a paradigm of bipersonal verbs where the undergoer is in the absolutive. The expression of an inanimate participant, such as *iyaʔ* ‘water’ is possible with both verbs, but it is optional, and the inanimate participant is not indexed on the verb, whether it is or is not overtly expressed. Both *na-ndrèhmé* and *na-ndrihmá* contain the same stem, which can approximately be written as NDRIHMA.²

There are four basic segmental paradigms encoding person and case. In (1) these four case-person paradigms are displayed, stripped of possible tones. ‘N’ refers to ‘new’—a category indexing a third person participant overtly introduced in the default position following the predicate. The symbol ‘~’ indicates predictable morphophonological alternation and the symbol ‘/’ an alternation which is predictable to a lesser degree, although it is clearly also phonologically conditioned, at least diachronically speaking.

(1) Case markers of monopersonal verbs

	Ergative	Absolutive	Pegative	Dative
1SG	-Ø	-ɥʔ	-u ~ -o	-uʔ ~ -oʔ
2SG	-Ø	-iʔ / -aʔ	-a / -i	-aʔ
3SG. N	-Ø	-i / -a	-u ~ -o	-u ~ -o
1PL/2PL	-Ø	-aʔ	-a / -i	-aʔ
3PL. N	-Ø	-i	-a / -i	-ɥ

Case roles always pertain to animate arguments. As is true of many case systems, semantics does not strongly predict which case is going to be used

² The form *na-ndrèhmé* contains an inflectional suffix *-i* (discussed in Section 2 below) which merges with the final *a* of the stem to produce *e*, a vowel which subsequently, by an entirely regular process, triggers harmony in the first vowel of the stem.

with a given verb, but there are some tendencies to be noted. Generally, the ergative is used when the animate actor carries out an action with a relatively high degree of impact (see a full list of examples in Section 4.2 below) and the pegative when the impact is weaker (e.g. *geʔdó* ‘s/he has it’, *nimindúʔ* ‘s/he saw it’); the absolutive and dative are used when the animate participant is in an undergoer role, with the absolutive (e.g. *daská* ‘s/he smells bad’, *namyahwí* ‘s/he worries’) signaling a somewhat higher degree of affectedness than the dative (e.g. *bašó* ‘s/he is nude’, *nahmyúʔ* ‘s/he is using it’).

There are two types of dipersonal verbs. In one of them the absolutive paradigm in (1) indexes the undergoer and the actor is indexed by other means, including tones. In the other type the undergoer is indexed by the dative paradigm in (1), except when a third person singular actor is involved, in which case it is the actor who is indexed, and for this person the pegative paradigm is used. Thus, the pegative (a case role for which I had to invent a neologism) is the mirror-image of the dative, referring to an actor involved in an event with a dative-like undergoer. Person marking for the participant not indexed by means of suffixes from either the pegative or dative paradigm is taken care of by other means, including tones. Examples (2-3), (taken from Wichmann 2010, to which the reader is referred for more information on case marking and indexing patterns) illustrate dipersonal verbs in the absolutive and the pegative, respectively.

- (2) *ni-guhpràʔ-á*
 PFV-kick-3SG.given>3SG.ABSOLUTIVE
 ‘S/he kicked him/her.’
- (3) *ni-raʔn-ú*
 PFV-meet-3SG.given.PEGATIVE>3SG
 ‘S/he met him/her.’

The absolutive, pegative, and dative each divide into a limited set of around five major and some minor tonal classes. These classes are possibly to some extent based on the different underlying tones of the stem plus a tone pertaining to the segmental suffix. Identifying underlying tones is difficult since verb stems normally do not occur uninflected, but the dative inflectional suffixes also serve to indicate possession on nouns, and in this case it can be directly observed how the tones of the free forms change when the noun becomes inflected. So a way to set up underlying tones for verbs

would be to observe internal tonal sandhi rules in paradigms of possessed nouns and then apply these rules to absolutive, pegative, and dative verbs, working backwards from inflected forms to underlying shapes. The limited number of tonal classes for absolutive, pegative, and dative verbs suggests that there are regular patterns (Wichmann 2006). Here I leave the question of the viability of this sort of analysis open since we are going to concentrate on the ergative paradigm, which is one that does *not* involve segmental encoding of case-person.

The universe of verbal inflection in Meʔpá is full of peculiar, language-specific facts with their own, internal logic in addition to very many irregularities. In this section we have only glanced superficially at this universe. Next, we enter into more detail, focusing exclusively on the ergative paradigm of monopersonal verbs. As we shall see, this sub-universe is already something of a Pandora's Box. But we will show that statistical regularities emerge from the apparent chaos. First the segmental morphology will be described. This is not crucial to the main points of this paper which concern tonal inflectional classes, a topic to which I turn in subsequent sections, but must necessarily be introduced in order for the reader to make sense of the paradigms that will be discussed.

3. Segmental morphology of the ergative inflection

The 149 verbs studied in this paper represent a convenience sample of verbs that I have elicited directly as a part of lexicographical efforts or have come across in the relatively large body of texts that I have collected. Before going on to listing these verbs in their entirety I will explain how the paradigms are structured. Let us start out observing the single paradigm in (4), which belongs to the most frequent tonal class.

In (4) I have chosen to present the imperfective paradigm. This aspect is indicated by *na-*. In the plural this morpheme is apparently replaced by *nu-*, but in a deeper analysis *nu-* can be regarded as composed of the same *na-* as in the singular and an element *u-* marking 'agentive plural'. In the potential aspect, marked by *ma-*, we see a parallel shift to *mu-* in the plural. In the perfective, however, which is marked by *ni-*, the prefix stays the same in the plural by a rule according to which an /i/ 'wins' when clashing with a

following /u/ –Meʔpá does not tolerate diphthongs. It is not necessary to show aspectual prefixes as we proceed in this exposition; instead I will be citing stems introduced by a hyphen to indicate that an aspectual prefix is required. All dynamic verbs require aspectual prefixes, whereas stative verbs, by definition, do not. For present purposes we will only be discussing dynamic verbs since the group of ergative verbs are all dynamic.

(4) Paradigm for HKAMA ‘hang up something’

1SG	na-hkamà	‘I am hanging it up’
2SG	na-ta-hkamà	‘you are hanging it up’
3SG.G	na-hkamé	‘s/he is hanging it up’
3SG.N	na-hkamà šabù	‘the man is hanging it up’
1PL.INCL	un-hkamà=luʔ	‘we (incl.) are hanging it up’
1PL.EXCL	un-hkamà=lòʔ	‘we (excl.) are hanging it up’
2PL	un-hkamà=làʔ	‘you (pl.) are hanging it up’
3PL.G	un-hkamé	‘they are hanging it up’
3PL.N	un-hkamà šabù	‘the men are hanging it up’

Another element marking the agentive is *ta-* in the second person singular. This *ta-* is the default member of a set of allomorphs that are largely predictable from the following rules, although there are more than ten forms in the sample that are completely irregular (not listed here):

- *t-* appears when the stem has an initial /r/, e.g. *-t-rahkwahyàʔ* ‘make wet (2SG)’;
- the prefix is replaced by zero when the stem has an initial /tr/ or /ta/ sequence, e.g. *-trigu* ‘shell, remove grains (2SG)’ and *-tagaši* ‘turn something around (2SG)’;
- *nda-* can provisionally be said to appear on monosyllabic stems not beginning in one of the consonants /b/, /d/, /j/, /n/, /nd/, /ʔ/ –when these occur *ta-* appears;³
- verbs in initial /ŋgV/ replace this syllable with /tV/ or /ndV/ in the second person, where the choice of /tV/ or /ndV/ is determined by the previous rule; when V is /i/ or /e/ the same vowel appears in the prefix, when it is /u/ or /o/ an /a/ appears in the prefix: *-ŋgiʔdù* / *-tì-ʔdù* ‘begin (1SG/2SG),⁴ *-ŋgembô* / *-nde-mbô* ‘gather together’ (1SG/2SG), *-ŋguwĩ* / *-nda-wĩ* ‘take out’ (1SG/2SG), *-ŋgošnú* / *-nda-šnú* ‘tell’ (1SG/2SG), *-ŋgongô* / *-nda-ŋgô* ‘carry’, *-ŋgušũʔ* / *-nda-šũʔ* ‘get up’

All members of the ergative group of verbs are agentive (similarly, all negatives are agentive; it is only among the absolutes and datives, where

³ The segmental inventory of Meʔpa is as follows (indicating IPA symbols in square brackets when they differ from the practical symbols used in this paper): /p/, /t/, /ç/ [tʃ], /č/ [tʃ̣], /k/, /ʔ/, /b/, /d/, /j/ [dʒ], /g/, /mb/, /nd/, /nj/ [ndʒ], /ŋg/, /s/, /š/ [ʃ], /h/, /m/, /n/, /ŋ/, /r/, /l/, /w/, /y/ [j], /i/, /e/, /a/, /u/, /o/, /i/ [ĩ], /e/ [ẽ], /a/ [ã], /u/ [ũ], /o/ [õ].

⁴ Some speakers have *-ta-giʔdù* in 2SG.

the animate participant is an undergoer, that we find cases where *ta-* and *u-* are absent, indicating patientivity). Agentivity is almost entirely a lexical category, only extremely few verbs can optionally take agentive marking or its absence (Wichmann 1996).

Moving further down the paradigm in (4) we find a contrast between *na-hkamé* ‘s/he (G) is hanging it up’ and *na-hkamà šabù* ‘the man (N) is hanging it up’ (here the noun *šabù* ‘man’ is provided to trigger the form which indicates that the indexed participant is new in the discourse –it is not a member of the paradigm as such). The same distinction is made for third person plural participants, cf. *un-hkamé* ‘they (G) are hanging it up’ vs *un-hkamàšabù* ‘the men (N) are hanging it up’. The category GIVEN is derived from NEW by segmental processes and in most cases it also has a different tone. Segmentally, GIVEN is marked by a suffix *-i*. This merges with a stem-final /a/ to yield *-e* (as in our example), something which triggers a harmonic lowering of the preceding stem vowel if this is /i/ or /u/ (e.g. *-riyaʔ* / *-reyéʔ* ‘take out (3N / 3G)’, *-rumà* / *-romě* ‘charge (money) (3N / 3G)’); it disappears when the stem-final vowel is /i/ (e.g. *-yašì* / *-yaší* ‘see (3N / 3G)’); and it replaces a stem-final /u/ or /o/ (e.g. *-yiʔtu* / *-yiʔti* ‘cut in two (3N / 3G)’), except when these are preceded by one of the consonants /ʔ, h, k, g/, in which case stem-final /u/ or /o/ plus suffixal *-i* yield a sequence /wi/ (*-hprigu* / *-hprigwí* ‘shake (3N / 3G)’). Tonally GIVEN is characterized by never being L or falling (HM, ML).

There are some tantalizing cases where *-i* is missing even though the NEW category is still marked tonally. The verbs in question, given here in their 1SG forms with class numbers referring to the tonal classes introduced in the next section, are in the following list.

- Class 4: *-waĩ* ‘think’, *-wahtǔʔ* ‘grab, hang up’, *-tihŋǎ* ‘empty’, *-tangǒ* ‘order’, *-škuʔdǎʔ* ‘put’, *-škahŋǎ* ‘scratch’, *-rahkwǎ* ‘throw’, *-mahwǐ* ‘throw down’, *-kawǐ* ‘singe’, *-hpiʔtǎ* ‘split (as wood)’;
- Class 5: *-yawǐʔ* ‘scratch’, *-šingǎ* ‘take off’, *-riʔi* ‘peal’, *-pahŋǎ* ‘move’, *-niŋǎ* ‘twist’, *-mahŋǎʔ* ‘move’, *-kámǎ* ‘bend over’, *-htangǎ* ‘turn something around’;
- Class 9: *-ŋgembô* ‘join’, *-ndĩ* ‘dig’, *-kungǎ* ‘chew, grind’, *-jĩ* ‘hear, listen’;
- Class 14: *-ŋgošnǔ* ‘tell’, *-ŋgongǒ* ‘carry’;
- Class 16: *-bǎ* ‘move’; Class 20: *-daʔ* ‘earn’;
- Class 22: *-rumbǎʔ* ‘work as a day-laborer’;
- Class 24: *-šiyǎ* ‘throw’.

The discerning reader will have noticed that several of the forms listed contain /i/ as the last stem vowel and may have wondered how it can be noticed in these cases that *-i* is missing –we remember that *-i* disappears for purely phonological reasons when the stem-final vowel is /i/. There is, however, an apparent pattern according to which missing *-i* pertains to all members of each tonal class in which the phenomenon occurs, so its absence can be inferred even when it cannot be directly observed. It is presently not certain whether the absence of *-i* for the marking of NEW has a grammatical motivation. Intriguingly, there is a pair of semantically close verbs that form a sort of minimal pair in the sense that one takes *-i* and the other does not: *hpiʔtǎ* ‘split (as wood)’ (Class 4, takes *-i*) vs *hpiʔtǎ* ‘crack (as stone)’ (Class 2, does not take *-i*). It is presently not clear whether forms not taking *-i* represent a special grammatical category with a certain, associated function, although this possibility is somewhat remote. Thus, it remains uncertain whether we are dealing with two larger inflectional classes of ergative verbs, those that take *-i* and those that do not.

The stem shapes for the three forms representing the category of plural speech act participants are identical, distinctions between first person singular inclusive and exclusive and second person plural being made by means of the enclitics =*luʔ*, =*lòʔ*, and =*làʔ*, respectively. The only exceptions to stems being identical are some cases where a monosyllabic stem with an initial glottal stop shifts the glottal stop to stem-final position in first person plural exclusive and second person plural (e.g. *-ʔšmì=luʔ* / *-šmìʔ=lòʔ* ‘sow, weave (1PL.INCL / 1PL.EXCL)’.⁵ For present purposes, then, it is sufficient to cite one stem for all without the enclitics, but bearing in mind that the enclitics contribute by making further distinctions within this sub-domain of person inflection.

⁵ Lexical stems only carry phonological tones on the last two syllables, but the elements treated here as enclitics carry their own tones. If they were to be treated as suffixes the general pattern of two tones per stem would be violated, which is the argument for not treating them as suffixes. On the other hand, as illustrated by the cases of a migrating glottal stop, their presence can incur phonological changes to the stem to which they attach, which is an argument for not treating them as independent words.

4. Tonal morphology of the ergative inflection

This tour through the paradigm in (4) in the previous section allows us to present a simplified version, shown in (5), where we leave out the aspectual prefix, take out the overt noun triggering the NEW category (while bearing in mind that it is necessary) and merge plural speech act participants into a single form, glossed as ‘SAP.PL’.

(5) Simplified paradigm for HKAMA ‘hang up something’

1SG	-hkamà
2SG	-ta-hkamà
3SG.G	-hkamé
3SG.N	-hkamà
SAP.PL	-u-hkamà
3PL.G	-u-hkamé
3PL.N	-u-hkamà

Having stripped down the paradigm to its essentials we can more easily focus on the tonal contrasts. Here the ‘odd men out’ are the 3SG.G and 3PL.G forms, which differ from the others by having a high tone. But these forms are also special in the sense that they encode an additional grammatical category. If it were not for the fact that the tones of the GIVEN cannot be systematically derived from the tones of the NEW forms they might be considered extraneous to the paradigm of person inflection. In any case, all other forms are homophonous, including their tones. As has been explained, the *ta-* and *u-* prefixes are portmanteau, encoding agentivity in addition to, respectively, second person singular and plural of all persons. The only non-portmanteau markers of person are the tones, and in the paradigm in (5) all the five fundamental categories –1SG, 2SG, 3SG.N, SAP.PL, 3PL.N– are neutralized. If it were always the case that tonal distinctions for these categories were neutralized it would not make sense to say that tones are a locus of person marking at all. However, the tonal inflectional class of which the paradigm of HKAMA is a representative, even if it is the one with the largest number of recorded members, is atypical in that it neutralizes what we have just described as the five fundamental categories.

We now turn to look at the tonal classes emerging by comparing the 149 paradigms that form the materials for this investigation. Before presenting the data it is pertinent to note that some of the five fundamental categories (1SG, 2SG, 3SG.N, SAP.PL, 3PL.N) are in fact always neutralized. Thus, 1SG

and SAP.PL always exhibit the same tones, as do 2SG and 3SG.N. This leaves 3PL.N as the third and last in a further reduced set of categories, which we can describe as first person (symbolized ‘1’) –stipulating, for the sake of practicality, that the SAP.PL form is fundamentally first person with second plural being derived–, non-first singular (‘NON1SG’), and third plural (‘3PL.N’). If all paradigms of the ergative are reduced to these three categories it turns out that there would be 17 classes. As mentioned, however, 3SG.G and 3PL.G are hard to isolate from the person-number paradigms. When these two categories are considered integral to the person-number paradigms there are 24 classes. These are displayed in Table 1. The first column assigns an arbitrary number to each class; the second column shows the number of members in each class; the third column gives a count of how many distinct tones (‘dist.’) are found in each paradigm; the subsequent columns shows the tones pertaining to each slot in the paradigm.

class	freq.	dist.	1	NON1SG	3PL.N	3SG.G	3PL.G
1	32	2	L	L	L	H	H
2	30	3	LM	M	M	H	H
3	28	3	L	M	M	H	H
4	11	4	LM	HM	ML	H	LM
5	8	3	ML	ML	ML	H	LM
6	5	2	L	L	L	LM	LM
7	4	3	L	M	L	H	M
8	4	4	L	M	L	H	LM
9	4	2	ML	ML	ML	H	H
10	3	3	L	L	L	H	LM
11	3	3	ML	LM	ML	MH	LM
12	3	5	LM	M	L	MH	H
13	2	2	LM	H	H	H	H
14	2	3	LM	ML	ML	LM	H
15	2	3	M	M	L	H	M
16	1	1	H	H	H	H	H
17	1	2	H	H	L	H	H
18	1	4	LM	LM	M	MH	H
19	1	3	M	M	L	M	H
20	1	2	M	M	L	M	M
21	1	3	L	M	M	MH	H
22	1	3	LM	LM	LM	MH	MH
23	1	4	LM	HM	HM	MH	H
24	1	3	LM	LM	ML	MH	LM

Table 1. Tonal classes of the Meʔpá ergative

Before continuing the analysis it should perhaps be argued why I have chosen to include among the paradigms some that have only one member. The reason

for this is that there is not an easily identifiable cut-off point among the lower-ranking paradigms at which one might distinguish an irregular paradigm from a regular one. In other words, a paradigm with two members is not by any non-arbitrary criterion regular, as opposed to one having only one member. Additionally, it may well be that some of the singletons in reality have more members which simply have not been recorded yet.

5. Negative evidence for purely morphological inflectional classes

I regard inflectional classes as being purely morphological when they are neither clearly phonologically nor clearly semantically or clearly grammatically motivated. The little recurring word ‘clearly’ is meant to leave the possibility open for using the term ‘pure morphological inflectional class’ even when there are traces of non-random motivations for class assignment. We will now be looking for such traces in the domains of phonology, semantics, and grammar.

5.1. Are there phonological motivations for class assignment?

This question can immediately be answered in the negative. The sheer number of tonal classes excludes the possibility of deriving the tonal paradigms from some systematic combination of underlying stem tones and inflectional ones. Moreover, the way that tones are distributed over paradigms is not suggestive of a single underlying tonal melody of the stem. For instance, compare the three paradigms in Table 2.

class	freq.	dist.	1	NON1SG	3PL.N	3SG.G	3PL.G
1	32	2	L	L	L	H	H
6	5	2	L	L	L	LM	LM
10	3	3	L	L	L	H	LM

Table 2. Selected subset of tonal classes of the Meʔpá ergative

For the data in Table 2 we might attempt to posit an underlyingly low tone in the stem to account for 1, NON1SG and 3PL.N. But how do we then go on to account for the patterns in the GIVEN forms? How can a low tone combined with certain tonal exponents of 3SG.G and 3PL.G sometimes yield H in both, sometimes LM in both, and sometimes H in one and LM in the other? This is but one example of how mind-boggling it would be to even begin to posit rules that would predict the patterns found.

5.2. Are there semantic motivations for class assignment?

In order to address this question we need to list the members of each class so as to be able to discern possible semantic relations. We again use the first person singular form as citation form. For the sake of completeness classes containing just one member are also listed. The Spanish glosses recorded (sometimes reflecting the regional variant) are also supplied in order to avoid ambiguities and loss of nuance in the translations.

- Class 1: *-hkamà* ‘hang / colgar’, *-hndraʔà* ‘throw / echar’, *-hndrigà* ‘throw / arrojar’, *-hpraʔà* ‘slice as with a machete / volar’, *-kogàraʔà* ‘make a fence / cercar, poner palos para cerca’, *-mbiyàʔ* ‘cry / llorar’, *-mihná* ‘fight / pelear’, *-ndahwà* ‘clean / limpiar’, *-ndaʔà* ‘borrow / pedir prestado’, *-ni* ‘do, make / hacer’, *-nù* ‘grind / moler, repasar’, *-ngewà* ‘measure, weigh / medir’, *-ngiʔdù* ‘begin / empezar’, *-nguhçù* ‘toast / tostar’, *-nguhwà* ‘sell / vender’, *-nguwàʔ* ‘complain / quejarse’, *-pahmigàʔ* ‘break / quebrar’, *-pahmigù* ‘crush / machacar, partir’, *-rahkwahyàʔ* ‘wetten/ remojar’, *-raʔà* ‘look (inside) / ver (para adentro), asomarse, quedarse viendo’, *-raʔçiʔ* ‘sneeze / estornudar’, *-rogàraʔà* ‘close the entrance to a house by tying poles together / cercar (amarrar los palos de la casa)’, *-rugà* ‘do manual labor, do a chore / manogrear, haciendo un quehacer’, *-ruhçi* ‘punch, crush / pilar, golpear, machacar’, *-ruwàʔ* ‘stab / acuchillar’, *-šnateʔyà* ‘lend / prestar’, *-šnatiʔkù* ‘distribute / repartir’, *-stiga* ‘put (multiple objects) / ponerlos’, *-tagašì* ‘turn something around / voltear’, *-taraʔà* ‘make a vow, pray / hacer promesa, rogar, publicar’, *-yašì* ‘see’, *-ʔgihyàʔ* ‘dry in the sun / asolearlo’;
- Class 2: *-hì* ‘boil / hervir’, *-çihpã* ‘squeeze out / exprimir’, *-çihtũ* ‘pinch / pellizcar’, *-çiʔgãʔ* ‘jump / brincar’, *-çwahũ* ‘put in vertical position (as several poles) / ponerlos (palos)’, *-hmidã* ‘make something burst, throw / reventarlo, tirarlo’, *-hpiʔtã* ‘crack (stone) / partir (piedra)’, *-hprigũ* ‘shake / sacudir’, *-htigũ* ‘throw down / tumbarlo para abajo’, *-kašì* ‘skin / despellejar’, *-kušì* ‘raise / alzar, levantar’, *-ndrihmã* ‘sprinkle / rociar’, *-ngidãʔ* ‘put / meter’, *-ngišì* ‘sweep / barrer’, *-ngundã* ‘dream / soñar’, *-pašì* ‘buck / reparar, corcovear’, *-ramìʔ* ‘whistle / chiflar’, *-raʔq̃* ‘look upwards / ver para arriba’, *-riyãʔ* ‘take out / sacar’, *-šigũ* ‘massage / sobar’, *-škuʔngãʔ* ‘push / empujar’, *-šmigũ* ‘pound, grind / martajar, remoler, repasar’, *-šmindũ* ‘scrub, smear / refregar, enjuagar’, *-šňã* ‘dance / bailar’, *-špibì* ‘shine / brillar’, *-tigũ* ‘let fall by accident / tumbar (algo se chispa de la mano, no es por la propia voluntad de uno)’, *-trigũ* ‘shell, remove grains / desgranar’, *-ʔgìʔ* ‘make (a plan) / hacer (plan)’, *-ʔyãʔ* ‘search for / buscar’;
- Class 3: *-yà* ‘take / agarrar’, *-yà* ‘cargar / to carry’, *-çigàʔ* ‘jump / brincar’, *-çigù* ‘put / poner’, *-çikà* ‘smoke / fumar’, *-dà* ‘shave oneself / rasurarse’, *-dàʔ* ‘throw, throw away / botar, tirar’, *-dù* ‘sow / sembrarlo’, *-gù* ‘pick up, gather / recoger’, *-hpaʔà* ‘throw inside / echar adentro’, *-migàʔ* ‘harvest (corn) / pizar’, *-nduhtàʔ* ‘spit / escupir’, *-nduʔwà* ‘laugh, smile / reir, sonreir’, *-niʔtù* ‘cut / cortar’, *-njaʔwà* ‘scream / gritar’, *-ngàʔmàʔ* ‘coger / to fuck’, *-riyàʔ* ‘mend / remendar’, *-rubù* ‘pull out / arrancar’, *-šìʔ* ‘throw up / vomitar’, *-šmì* ‘sow, weave / coser, tejer’, *-šnà* ‘give / dar’;

-*šnaʔši* ‘hand in / entregar’, -*šnà* ‘dance / bailar’, -*spà* ‘cut weeds or underbrush in a collaborative fashion / fajinear’, -*tà* ‘talk, speak / hablar’, -*tamà* ‘hang up / tender’, -*yahùʔ* ‘work’, -*ʔgiʔ* ‘put (multiple objects) / ponerlos’;

- Class 4: -*ɕadũ* ‘step on, crush / pisar, machucar’, -*hpiʔtã* ‘split / rajar’, -*kawĩ* ‘singe / chamuscar’, -*mahwĩ* ‘knock down / tumbar’, -*rahkwã* ‘throw / echar’, -*škahŋgã* ‘scratch / rascar’, -*škuʔdãʔ* ‘poner / to put’, -*taŋgõ* ‘order / mandar’, -*tihŋgã* ‘empty / vaciar’, -*wahtũʔ* ‘grab, hang up / agarrar, tender’, -*wai* ‘think / pensar’;
- Class 5: -*htaŋgã* ‘turn over / voltear’, -*kámã* ‘bend over / agacharse’, -*mahŋgãʔ* ‘move / mover’, -*niŋgã* ‘twist / torcer’, -*pahŋgã* ‘move / mover’, -*riʔi* ‘remove leaves, peel / deshojar, pelar’, -*šingã* ‘take off / quitar’, -*yawĩʔ* ‘scratch / rascar’;
- Class 6: -*ɕadomà* ‘mend / remendar’, -*máhq̄* ‘make wet / mojar’, -*ŋgoʔmà* ‘stick / pegarse’, -*rúhtù* ‘cut up / trozar, cortar’, -*rumà* ‘charge (money) / cobrar’;
- Class 7: -*ɕwadàʔ* ‘insert / meter (perforar)’, -*gùèʔdà* ‘carry / cargar’, -*hyà* ‘wash / lavar’, -*šum̀bùʔ* ‘roast / asar’;
- Class 8: -*škahmà* ‘touch / tentar’, -*škuʔtù* ‘break / romper’, -*yamà* ‘spy on / espiar’, -*yĩʔtù* ‘cut, cut up / cortar, trozar’;
- Class 9: -*jũ* ‘oír, escuchar / to hear, listen’, -*kun̄gã* ‘chew, grind / mascar, remoler’, -*ndĩ* ‘dig / escarbar’, -*ŋgembô* ‘join / juntar’;
- Class 10: -*kaʔò* ‘hide / esconder’, -*kušpà* ‘to suck / chupar’, -*ŋguhçù* ‘fry / asar’;
- Class 11: -*ndihtĩʔ* ‘fan / ventear’, -*raʔwĩ* ‘choose, take out / escoger, sacar’, -*šidĩʔ* ‘throw / tirar’;
- Class 12: -*riʔyãʔ* ‘add to / echarle’, -*šiʔyãʔ* ‘throw / echar’, -*štaʔmãʔ* ‘arrange, put (multiple objects) / colocarlos, ponerlos’;
- Class 13: -*ɕĩ* ‘play / jugar’, -*ŋgušũʔ* ‘get up / levantarse’;
- Class 14: -*ŋgongõ* ‘carry / acarrear’, -*ŋgošnũ* ‘count / contar’;
- Class 15: -*gùrìgu* ‘receive (multiple objects) / recibirlos’, -*wèʔka* ‘confess / confesar’;
- Class 16: -*bq̄* ‘move / menear, mover’;
- Class 17: -*ŋgišpá* ‘squash (as plastic bottle) / pachacar’;
- Class 18: -*rigũ* ‘cut / cortar’;
- Class 19: -*hndùwaʔ* ‘cheat / engañar’;
- Class 20: -*daʔ* ‘win / ganar’;
- Class 21: -*štaʔmàʔ* ‘put up / poner arriba’;
- Class 22: -*rumbãʔ* ‘work as a day-laborer’;
- Class 23: -*ŋguwĩ* ‘take out / sacar’;
- Class 24: -*šiyã* ‘throw / tirar’.

If there are semantic commonalities to be found among the members in each class we would expect these to most easily be found in the smaller classes, but already here we run into problems. Do such items as ‘receive (multiple objects)’, ‘confess’ (Class 15) or ‘carry’, ‘count’ (Class 14) or ‘play’, ‘get up’ (Class 13) have something in common? Hardly. The problems only increase when we move further up in the ranks. For instance, Class 9, with four members, includes ‘hear, listen’, ‘chew, grind’, ‘dig’, ‘join’. We can continue to cite the members of still larger classes but it would only serve to show that none of them has any internal semantic coherence. It cannot be excluded that some members are attracted to a certain class by local affinities. For instance, several members in Class 3 refer to acts of putting / throwing / giving or opposed types of acts such as taking / picking / pulling, but these meanings far from exhaust the semantic range of Class 3 members, and similar meanings are also found in other classes. Thus, it is fair to say that the various classes are not in any obvious, systematic ways based on semantic criteria.

5. 3. Are there grammatical motivations for class assignment?

The group as a whole assigns ergative case to the single animate participant, who is usually a voluntary agent in control of the action, which, when involving an inanimate undergoer, usually implies a relatively high degree of effect. The question is whether individual classes encode additional distinctions. The best candidate seems to be a distinction associated with a missing *-i* and certain tonal inflectional classes (4, 5, 9, 14, 16, 20, 22, 24), the phenomenon mentioned in Section 2 above. Much as I would like to, I have not been able to pinpoint some function that would tie together the members of these eight classes as opposed to the others.⁶

⁶ For the Miꞑphà: (Malinaltepec) variant Suárez (1983:139) notes that approximately one half of the verbs in his ‘D7’ group –the one that I label as the ergative– take *-i*. For Meꞑpá (Azoyú) the proportion is approximately four fifths. Some examples of Miꞑphà: verbs that do not take *-i* while their Meꞑpá cognates do, are *-njàꞑwà* ‘scream / gritar’, *-čìpa* ‘knead / amasar’, *-čìthú:* ‘arañar / to scratch’, *-ndùꞑwà* ‘laugh / reirse’, *-nò* ‘grind / moler’, *-rúbu* ‘pull out / arrancar’, and *-šùꞑmbú* ‘roast, grill / asar’ (cited from Suárez 1983: 139-140 with conversion of the orthography to the one used in this paper and addition of English glosses). The fact that *-i* can ‘come and go’ between different Tlapanec variants would not be expected if its presence/absence were associated with specific grammatical functions.

6. Positive evidence for purely morphological inflectional classes

The conclusion to draw from the previous section is that there are no positive criteria for membership in the different tonal classes. This constitutes negative evidence in favor of the existence of purely morphologically motivated inflectional classes. Now we will consider different statistical aspects of the 24 classes based on the data in Table 1.

The first question to be addressed is whether there is a correlation between the size of classes and the number of distinct tones that occur in the paradigms (the ‘dist.’ column). The two quantities turn out to be uncorrelated ($r = -0.055$, $p = 0.800$). The number of distinct tones varies from one to five, which are also respectively the theoretical minimum and maximum, and the average is 2.9. If there was a positive correlation to be found we might take that as evidence for a drift towards maximal distinctiveness among the five cells and if there was a negative correlation we could assume the existence of an opposed tendency to neutralize distinctions. These inferences rely on the assumption that the larger classes are more representative of the direction in which the language is moving (and furthermore on the assumption that a language constantly changes, but this should hardly be controversial). Rather than a drift towards either maximal or minimal distinctiveness, there seems to be a tendency to keep an intermediate number of distinctions, with roughly two pairs of cells showing tonal homophony on average.

The second test I wish to make is to see whether a class being large also means that it is more similar to other classes overall. If so, it would imply that there is a drift such that tones in smaller classes are changed to conform to those of larger classes. To investigate this possibility the number of similar tones in corresponding cells of all possible pairs of paradigms was computed. For instance, class 1 and 2 have two identical cells so they get a similarity of 2; class 1 and 3 have three identical cells so they get a similarity of 3; and so on. From the resulting matrix the sum of each column representing a class and its similarities to the 23 other classes is computed, and these measures of overall similarity between each and every other class are correlated with the size of the classes (since the experiment is easily replicated using the data in 4 I will not display a 24 x 24 table of numbers here). The result of this exercise is a relatively solid Spearman rank correlation which is statistically significant ($\rho = 0.446$, $p = 0.029$). In other

words, there is, indeed, at least a weak tendency for larger classes to overall be more similar to the other classes than smaller ones. I interpret this result as telling us that when tones shift they tend to shift towards tones in corresponding cells of paradigms with more members.

The third test concerns the statistical distribution of class sizes. It is immediately apparent from Table 1 that there are a few larger classes, some more medium-sized ones and many small ones. This sort of distribution resembles what linguists readily recognize as a Zipfian distribution, which is more generally known as a power-law distribution. Such a distribution will show a close fit to a straight line when the quantities are plotted against their ranks on a log-log diagram. This is done for our data in Figure 1. The best fit of a power function to the data is $N = 63.426 \cdot R^{-1.34}$, where N is the size and R the rank of a class. The fit is extremely good and highly significant ($r = -0.973$, $p < 1e^{-14}$).⁷

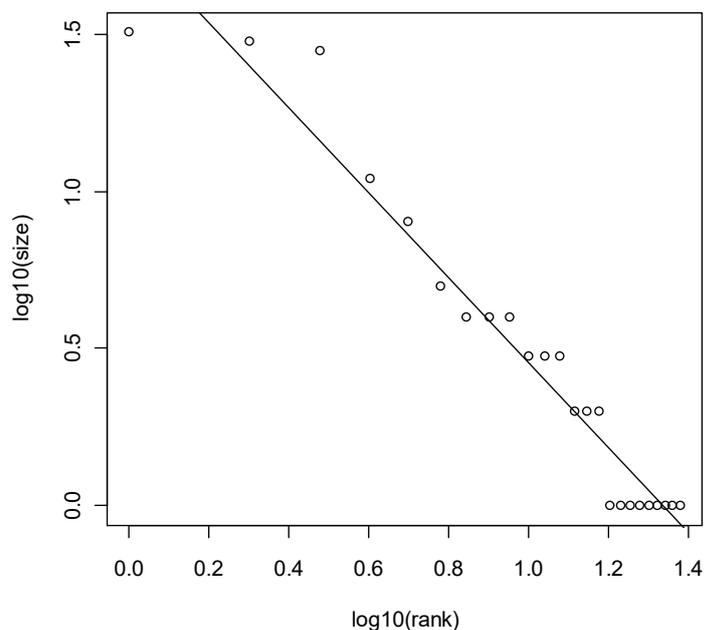


Fig. 1. A log-log plot of sizes of classes as a function of the ranks of their sizes

In spite of the neat distribution seen in Figure 1 there are outliers among the dots representing the largest sizes. The fit would increase if either the first or the third dot from the left were taken out. This is interesting in light of the

⁷ If the classes characterized by a missing *-i* in the NEW category are removed from the distribution the fit to a power-law is still very good, only slightly lowered to $r = -0.967$.

fact that class 2 and 3 are so similar that we can easily imagine them drifting towards one another –the only difference is the tone in the first cell. As a matter of fact, I did record at least two clear instances where two speakers disagreed about the class membership. As the first person stem of $\text{CIGA}ʔ$ ‘jump’ one speaker gave $-\text{çigà}ʔ$ whereas another gave $-\text{çigǎ}ʔ$. Similarly the first person of SDA ‘dance’ was recorded as $-\text{snà}$ and $-\text{snǎ}$ from different speakers. For many verbs I elicited paradigms from more than one speaker, but did not do this systematically for all verbs, so the two instances noted here are likely to be symptomatic of a larger general confusion between the two classes, either among speakers or in my data. That is, either there is a single class with a difference in the realization of the first cell of the paradigm across speakers, in which case I have assigned verbs to different classes depending on which speaker I was working with, or there are two classes with agreement across speakers concerning some of its members and disagreement concerning others. In the latter case this would be symptomatic of a merger of classes in the direction of a better fit with the power-law. In the former case we would already have a better fit. Merging classes 2 and 3 into a single class with 58 members increases the fit of the power-law to a remarkable $r = -0.987$.

Taken together, the second and the third test show what can be taken as positive evidence for pure inflectional classes. The fact that their rank-size distribution obeys a power-law supports the hypothesis that they are driven by stochastic processes where larger classes work as attractors on smaller ones. But we should caution that such a distribution cannot on its own prove such a hypothesis since power-law distributions can arise in many different ways, as discussed in the next section. For instance, we can imagine that a semantic motivation for different verb classes can lead to a similar distribution if some of the semantic situations or categories encoded are very broad or frequent, several less broad or frequent, and many are quite specialized.⁸

⁸ Future work will be directed at also looking at tonal classes of each of the other three grammatically defined verb classes –those taking the absolutive, dative and pegative– to see how their type frequencies are distributed and to find possible arguments for or against phonological motivations for these tonal classes. A preliminary survey has shown that the frequencies of the tonal classes for the dative definitely do not show a power-law distribution; and, as argued in Wichmann (2006), they also appear to be phonologically motivated, being predictable from the tone of the penultimate tone of the lexical stem. For the absolutive and the pegative the distributions are more along the lines that of the ergative, but these distributions need to be determined more precisely and the tonal paradigms must be studied in more detail before any conclusions can be drawn.

7. Discussion and conclusion

Power-laws are ubiquitous in the world that surrounds us. They have been found to describe the size-distribution of natural phenomena like, for instance, the abundance of biological taxa, the size of earthquakes or the activity of genes. They also recur in social phenomena like city sizes, personal income, citation patterns, and different kinds of social networks.⁹ Finally, linguists are familiar with Zipf's law, which predicts that word frequencies are inversely proportional to their Rank –yet another power-law phenomenon, and also among one of the first to have been identified.

The steadily growing literature abounds with mathematical models that describe how such distributions can come about, but for the present purposes it is sufficient to observe that the phenomenon can arise by a process known as preferential attachment whereby centrally connected individuals in a network will tend to increase their connectivity. In the preceding section it was suggested that a similar sort of process might operate within the group of ergative verbs in Meʔpá, with larger classes working as attractors on the members of smaller ones. While it is hardly controversial that languages are in constant flux –a necessary condition for the power-law distribution of verb classes to be an outcome of historical processes– we have also seen a glimpse of evidence for this through some verbs whose tonal shapes vary from speaker to speaker, as noted in the preceding section.

The many classes of Meʔpá ergative verbs is not exactly the sort of thing I had hoped to be confronted with when, in the early 1990's, I collected the data on which this paper is based. The present writer, then a youthful field worker, was hoping to quickly be able to write a tidy reference grammar of the language, but was brought to some degree of despair. Having acquired more insights into stochastic processes it has now become clear to me that the sort of size distribution observed is exactly what one would expect from a rich morphological system which is as purely morphological as such a system can possibly get, without 'contamination' from phonology, semantics, and functionally motivated grammatical distinctions.

⁹ In Wichmann (2005b), I pointed out that language family sizes are also distributed according to a power-law, and in this paper included some straightforward discussion of the phenomenon which may be consulted for further background (the paper also includes some potentially useful core references).

Acknowledgements

I would like to express my appreciation for comments by Bernard Comrie on an earlier version of this paper. I am also grateful to the editors for their comments. The research was funded by an ERC Advanced Grant (MesAndLin(g)k, Proj. No. 295918) and by a subsidy of the Russian Government to support the Program of Competitive Development of Kazan Federal University.

References

- Suárez, Jorge A. 1983. *La lengua tlapaneca de Malinaltepec*. Mexico City: Universidad Nacional Autónoma de México.
- Wichmann, Søren. 1996. The degrammaticalization of agentivity in Tlapanec. *Content, Expression and Structure: Studies in Danish Functional Grammar*. Studies in Languages Companion Series, E. Engberg-Pedersen, M. Fortescue, P. Harder, L. Heltoft & L. Falster Jakobsen (eds), 343-360. Amsterdam/Philadelphia: John Benjamins.
- Wichmann, Søren. 2005a. Tlapanec cases. *Conference on Otomanguean and Oaxacan Languages, March 19-21, 2004*. Report 13, Survey of California and Other Indian Languages, R. Beam de Azcona & M. Paster (eds), 133-145. Berkeley: University of California at Berkeley.
- Wichmann, Søren. 2005b. On the power-law distribution of language family sizes. *Journal of Linguistics* 41: 117-131.
- Wichmann, Søren. 2006. Sandhi tonal interno en la morfología verbal tlapaneca. *Memorias del VIII Encuentro Internacional de Lingüística en el Noroeste*, vol. 2, R.M. Ortiz Ciscomani (ed.), 337-355. Hermosillo: Editorial UniSon.
- Wichmann, Søren. 2007. The reference-tracking system of Tlapanec: between obviation and switch reference. *Studies in Language* 31.4: 801-827.
- Wichmann, Søren. 2009. Case relations in a head-marking language: verb-marked cases in Tlapanec. *The Handbook of Case*, A. Malchukov & A. Spencer (eds), 797-807. Oxford: Oxford University Press.
- Wichmann, Søren. 2010. Ditransitive constructions in Tlapanec. *Studies in Ditransitive Constructions: A Comparative Handbook*, A. Malchukov, Andrej, M. Haspelmath & B. Comrie (eds), 651-677. Berlin/New York: Mouton de Gruyter.

