VeLePa: a Verbal Lexicon of Pame

Borja Herce
University of Zurich

Abstract

This paper presents VeLePa, an inflected verbal lexicon of Central Pame (pbs, cent2154), an Otomanguean language from Mexico. This resource contains 12528 words in phonological form representing the complete inflectional paradigms of 216 verbs, supplemented with use frequencies. Inflected lexicons of non-WEIRD underresourced languages are urgently needed to expand digital capacities in these languages (e.g. in NLP). VeLePa contributes to this, and does so with data from a language which is morphologically extraordinary, with unusually high levels of irregularity and multiple conjugations at various loci within the word: prefixes, stems, tone, and suffixes constitute different albeit interrelated subsystems.

1 Introduction

Central Pame is an indigenous Mesoamerican language spoken by around 5000 people in and around Santa María Acapulco (San Luis Potosí, Mexico). The language is still acquired as a first language by children in various communities, but is endangered by the expansion of Spanish. The language lacks a standard written form, and extant documentation (e.g. Gibson & Bartholomew, 1979; Hurch; 2022) is insufficient, undigitized, and computationally largely unusable.

The language, however, like others in its family (e.g. Chichimec, see Palancar & Avelino, 2019; Herce, 2022) is a treasure trove of morphological complexity, due to the combination of the following two traits:

- Very high levels of irregularity, with many small inflection classes, many uniquely-behaving verbs, and a lot of suppletion.
- A morphological realization of subject and tense information which is distributed along the word into multiple inflectional layers: prefixes, tone, stem, and suffixes.

These properties make the system highly interesting and challenging to theoretical morphology as well as to NLP. Adding this language to databases like Unimorph (see McCarthy et al., 2020) and to morphological reinflection tasks would make these more representative of overall human language diversity and its limits.

2 Building VeLePa

To build an inflected lexicon of Central Pame verbs the first thing we need is language documentation. Although some inflectional paradigms were collected by SIL missionaries around 70 years ago (Gibson, 1950), these are insufficient in number and are hardly usable computationally due to inconsistencies.

Over the last four years, I have been documenting the language together with native speakers, mostly through the elicitation of inflected forms. For their orthographic transcription I adopt a phonemic approach, whereby only contrastive sounds are represented with different characters. International Phonetic Alphabet conventions are followed, as in the aforementioned previous work on the language. I thus avoid the problems of a Spanish-based orthography that is occasionally used to write the language locally and which does not represent features like vowel nasality, tone, consonant length, and the contrast between an mid-open and mid-closed front vowels.

The database (VeLePa) that is presented in this paper contains therefore the complete paradigms of a large number of verbs in phonological form. Every single one of the 12528 inflected forms that VeLePa contains (all 58 forms from 216 verbs) has been independently elicited (i.e. never extrapolated from other forms, as is often the case of these recources) and checked multiple times to avoid mistakes and inconsistencies (e.g. in the treatment of synonymous inflected forms like *dived*～*dove*). This is needed, first, because the language demands it. Most of the words in VeLePa (74%) have different forms, and syncretism (i.e. morphological whole-word identity) is never the result of different values being systematically the same across all lemmas as in other languages (e.g. English *do* INF, *do* 1SG.PRS, *do* 2SG.PRS, *do* 1PL.PRS, *do* 2PL.PRS, *do* 3PL.PRS).

Proceedings of the 21st SIGMORPHON workshop on Computational Morphology, Phonology, and Phonetics, pages 1–6

Secondly, given the large degree of irregularity in the language, the linguist can almost never be sure to predict correctly one form of a verb from another. Eliciting every single form prevents underestimating complexity. At the same time, however, because VeLePa has been built with computational analysis in mind, cross-speaker and intra-speaker variability and free variation had to be ironed out in a way that this does not lead to an *over*estimation of morphological complexity. Although crucial, these types of quality controls are not always discussed and implemented in the compilation of inflected lexicons, particularly those from indigenous languages, as these tend to be produced by documentary linguists for whom the computational use of these resources is not a priority.

Given the absence of a standard of the language, and the unsuitability[1] of the orthography generally in use in the community, forms are represented in VeLePa in phonological transcription. Tones (High [H], Low [L], and Falling [F]) are indicated immediately after the (lowest) vowel of the syllable where they occur. Consonant gemination is indicated through a doubling of the corresponding consonant. To facilitate analysis, segmentations of prefix and stem have been included (indicated by "-"), as well as zero prefixes (indicated by "0"). These can be deleted if morphological decomposition is not needed. Other transcription choices are IPA-compliant. Typical forms are hence to-hoHʔo, 0-mbãLnʔ, laH-ppo, la-hõFlʔ, etc. or from a single verb la-nõH , ta-nõHn, ki-ŋõHik, 0-nõH, etc.

Every inflected form is tagged for its lemma (e.g. 'play') and morphosyntactic values (e.g. 1SG.PRS). As a further feature of interest to computational morphologists, for example those interested in the Paradigm Cell Filling Problem in a naturalistic setting, (see Ackerman et al., 2009; Blevins et al., 2017), I also provide a use frequency estimate of the different lemmas (see Figure 1, frequency estimated in number of tokens per million words) and morphosyntactic values (see Figure 2, frequency estimated as proportion of verbal tokens). These were derived from the frequency of forms in extant Central Pame texts (see Gibson et al, 1963; Gibson, 1966; Hurch 2022), and supplemented with subjective frequency estimates from native speakers (see Carrol, 1971) due to the small size of the available corpus (only 1171 verbal tokens) and its unbalanced thematic and genre composition.

[1] While these are phonemic in the language, neither tone nor vowel nasalization nor consonant gemination are consistently represented in the traditional orthography.

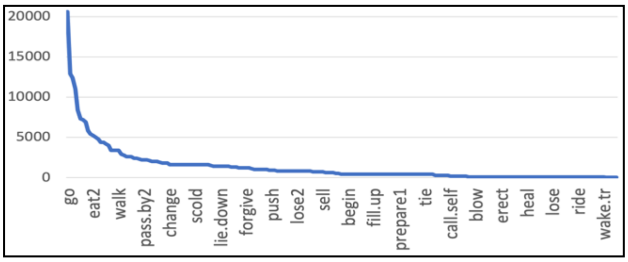

Figure 1: Frequency rankings of lemmas

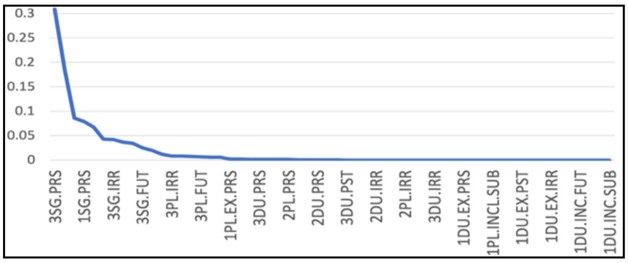

Figure 2: Frequency rankings of values

3 Analysis of system complexity

On the basis of VeLePa, freely available online at `https://osf.io/xhyzm/?view_only=763f1c043e3f4c3787d0c93226e8b817`, I analyze the morphological complexity and the predictability of the inflectional system as per the Paradigm Cell Filling Problem (see Ackerman et al., 2009). As mentioned in the introduction, one of the key idiosyncratic features of the language is the relative independence of prefixal, suffixal, tonal, and stem morphology. These four layers are analyzed separately below, through the following software:

- Qumín (Beniamine, 2018), for the automatic extraction of morphological alternations, and for the calculation of Information-Theoretic measures (e.g. conditional entropy of one form given another).
- Principal Parts Analyzer (Stump & Finkel, 2013), for the calculation of Set-Theoretic measures like the number of principal parts (i.e. the lowest number of forms required to predict the complete paradigm).

3.1 Prefixes

Despite their exhuberant allomorphy and the presence of stem-initial alternations, prefixes are straightforward to segment from stems. As the exemplary forms in Section 2 suggest, the prefix is the most changeable part of the word, and setting aside cases of zero-prefixed forms, corresponds generally to the first syllable of the word. Given this identification of prefixes, Pame verbs classify into 22 different inflectional classes, with a few comparatively frequent ones (see Table 1), and a long tail of (12) verbs which are prefixally unlike any other in the database.

type freq.	85	51	24	10	9	6	5	5
1SG.PRS	la	to	ti	la	la	ti	la	to
1DU.EX.PRS	ta	to	ti	ta	ta	ti	ta	to
1DU.INC.PR	ta	to	ti	ta	ta	ti	ta	to
1PL.EX.PRS	ta	to	ti	Ø	ta	ti	wa	to
1PL.INC.PRS	ta	to	ti	Ø	ta	ti	wa	to
2SG.PRS	ki	to	ti	ki	ki	ti	ki	la
2DU.PRS	ki	to	ti	ta	ki	ti	ta	la
2PL.PRS	ki	to	ti	Ø	ki	ti	wa	la
3SG.PRS	wa	lo	li	Ø	Ø	li	Ø	wa
3DU.PRS	wa	lo	li	Ø	Ø	li	Ø	wa
3PL.PRS	Ø	wa	ti	Ø	Ø	li	wa	wa

Table 1: Present prefixes of the 8 largest classes

As Table 1 shows,11 values of person-number are distinguished in the language, over 6 values of tense-aspect-mood. Due to the incompatibility of 1st and 3rd persons with the imperative mood, 58 values/cells exist in the Pame verb's paradigm. These fall into 39 areas of mutual interpredictability (see Table 2). These are those areas where the content of one cell (e.g. the 1PL.EX.PRS) allows to to predict that of another (e.g. 1PL.INC.PRS) and *vice versa*. In Pame this tends to mean their forms are always the same (e.g. ta/ta, to/to, ti/ti., Ø/Ø, or wa/wa in Table 1).

	PRS	PST	IRR	SUB	FUT	IMP
1SG	1	9	16	24	32	-
1DU.EX	2		17	25		-
1DU.INC						-
1PL.EX	3	10	18	26	33	-
1PL.INC						-
2SG	4	11	19	27	34	37
2DU	5	12	20	28	32	38
2PL	6	13	21	29	33	39
3SG	7	14	22	30	35	-
3DU						-
3PL	8	15	23	31	36	-

Table 2: Prefix interpredictability areas

The average conditional entropy (i.e. a measure of the uncertainty involved in predicting one form from another) is 0.52 bits. On a different metric of complexity, 5 static principal parts are needed to predict the entire paradigm. These speak of the complexity of prefixal inflection in Central Pame, which is, however, lower than that of the other inflectional layers/subsystems in the language that will be presented in the next sections.

3.2 Stems

While all Pame verbs show prefixal and suffixal inflection, not all (96.3%) display stem alternation. Barring cases of suppletion, which occurs in twelve verbs, generally with different roots in SG/DU and PL, most of the morphological action in stems occurs on their consonantal onset. Sometimes, particularly in the 3PL across tenses, it involves the addition of segments, some other times it involves gemination, sometimes segmental changes, etc. These occur with somewhat recurrent distributions in the paradigm (see a summary of the largest classes in Table 3).

type freq.	16	13	6	5	5	5	5	5
1SG.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
1DU.EX.PR	pp	ʔ	ʔu	h	kk	pp	tt	tt
1DU.INC.P	pp	ʔ	ʔu	h	kk	pp	tt	tt
1PL.EX.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
1PL.INC.PR	pp	ʔ	ʔu	h	kk	pp	tt	tt
2SG.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
2DU.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
2PL.PRS	ppy	ʔy	ʔu	h	kky	pp	kky	kky
3SG.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
3DU.PRS	pp	ʔ	ʔu	h	kk	pp	tt	tt
3PL.PRS	b	lʔ	tʔ	th	kh	pp	lh	lʔ
1SG.PST	w	ʔu	ʔu	h	ku	pp	t	t

Table 3: Present stem onsets of the 8 largest classes

Given the regularities in the distribution over values of different alternations, the 58 cells of the Pame verb paradigm are grouped into 29 interpredictability areas (see Table 4). The average conditional entropies between them is 0.63 bits, and 6 principal parts are minimally needed to be able to predict the complete stem paradigm without uncertainty.

	PRS	PST	IRR	SUB	FUT	IMP
1SG	1	7	13		22	-
1DU.EX						
1DU.INC						
1PL.EX	2	8	14	19	23	
1PL.INC						
2SG	3	9	15		24	28
2DU						
2PL	4	10	16	20	25	29
3SG	5	11	17	21	26	-
3DU						
3PL	6	12	18		27	

Table 4: Stem interpredictability areas

3.3 Tones

Tone (high, falling, or low) occurs in Pame in the stressed syllable, which can be either the final one (i.e. the root), or the penultimate (i.e. the prefix). Tone and stress are further intertwined in the language in that only the high tone occurs when the stressed syllable is the penultimate. The result is that only 4 tone-stress profiles are possible in any given word.

While all or most Pame verbs are inflectable in the other morphological layers, tone is different in that most verbs (66.2%) have a single tone across the paradigm (see the 4 largest classes in Table 5). Despite this, the PCFP is a considerable challenge because there is no way to predict, from the tonal value of a given form, whether this same tone will be found across the paradigm or in specific domains only, of which 19 exist (see Table 6).

type freq.	52	47	25	18	8	5	5	4
3SG.PRS	-L	-H	H-	-F	-L	H-	-H	-H
3DU.PRS	-L	-H	H-	-F	-L	H-	-H	-H
3PL.PRS	-L	-H	H-	-F	-L	-H	-H	-H
1SG.PST	-L	-H	H-	-F	-L	-H	-L	-H
1DU.EX.PST	-L	-H	H-	-F	-L	-H	-L	-H
1DU.INC.PS	-L	-H	H-	-F	-L	-H	-L	-H
1PL.EX.PST	-L	-H	H-	-F	-L	-H	-L	-H
1PL.INC.PST	-L	-H	H-	-F	-L	-H	-L	-H
2SG.PST	-L	-H	H-	-F	-F	-H	-F	-L
2DU.PST	-L	-H	H-	-F	-F	-H	-F	-L
2PL.PST	-L	-H	H-	-F	-F	-H	-F	-L

Table 5: Tones of the 8 largest classes

<table>
<tr><th></th><th>PRS</th><th>PST</th><th>IRR</th><th>SUB</th><th>FUT</th><th>IMP</th></tr>
<tr><td>1SG</td><td rowspan="3">1</td><td rowspan="3">6</td><td rowspan="3" colspan="2">10</td><td rowspan="3">16</td><td rowspan="5">-</td></tr>
<tr><td>1DU.EX</td></tr>
<tr><td>1DU.INC</td></tr>
<tr><td>1PL.EX</td><td rowspan="2">2</td><td rowspan="2">7</td><td rowspan="2" colspan="2">11</td><td rowspan="2">17</td></tr>
<tr><td>1PL.INC</td></tr>
<tr><td>2SG</td><td rowspan="2">3</td><td rowspan="2">8</td><td rowspan="2" colspan="2">12</td><td rowspan="2">18</td><td rowspan="2">12</td></tr>
<tr><td>2DU</td></tr>
<tr><td>2PL</td><td>4</td><td>9</td><td colspan="2">13</td><td>19</td><td>13</td></tr>
<tr><td>3SG</td><td rowspan="2">1</td><td rowspan="2">6</td><td rowspan="2" colspan="3">14</td><td rowspan="3">-</td></tr>
<tr><td>3DU</td></tr>
<tr><td>3PL</td><td>5</td><td>7</td><td colspan="3">15</td></tr>
</table>

Table 6: Tone interpredictability areas

Despite the small number of possible values of tone, the average conditional entropy between these domains is 1.01, and one would need minimally 7 principal parts to be able to predict with certainty the tone of every inflected form. These values are the highest among all four inflectional layers.

3.4 Suffixes

While prefixes, stems, and tones encode, often redundantly, different values of subject person-number, and tense-aspect-mood, suffixes tend to encode person-number almost exclusively. Pame suffixes are always non-syllabic, attaching as a syllable coda when the stem finishes in a vowel (e.g. kowwaL +i > kowwaLi; kowwaL +nʔ > kowwaLnʔ) but modifying the stem ending when the root already has a coda (e.g. toŋgoãHn +i > toŋgoãHiŋ, toŋgoãHn +nʔ > toŋgoãHnʔ). This gives rise to unpredictability in that, given a suffixed form (e.g. one which contains an underlying suffix -nʔ), it cannot be known what the unsuffixed form is (e.g. Ø vs -n in the verbs above).

Alongside this source of maybe "superficial" unpredictability, suffixes also change from verb to verb. As the forms in Table 7 show, some have a 2DU suffix -k while others do not, and some have a 3PL suffix -t while others do not. Mainly these two sources of unpredictability combine to generate a PCFP challenge comparable to the other inflectional layers, with 14 areas of interpredictability (see Table 8), 0.62 bits of average conditional entropy, and 6 static principal parts.

type freq.	50	27	22	8	8	6	6	6
1SG.PRS	Ø	ŋ	ʔ	n	Ø	ʔ	ŋ	t
1DU.EX.PR	mʔ	mʔ	mʔ	nʔ	mʔ	mʔ	mʔ	nʔ
1DU.INC.PR	Ø	Ø	ʔ	ŋ	Ø	ʔ	Ø	t
1PL.EX.PRS	nʔ	nʔ	nʔ	nʔ	nʔ	nʔ	nʔ	nʔ
1PL.INC.PR	n	n	n	n	n	n	n	n
2SG.PRS	Ø	ŋ	ʔ	n	Ø	ʔ	ŋ	t
2DU.PRS	Ø	Ø	ʔ	ŋ	k	ʔk	Ø	t
2PL.PRS	n	n	nʔ	n	n	nʔ	n	n
3SG.PRS	Ø	ŋ	ʔ	n	Ø	ʔ	ŋ	t
3DU.PRS	Ø	Ø	ʔ	ŋ	Ø	ʔ	Ø	t
3PL.PRS	Ø	ŋ	ʔ	n	t	ʔ	nt	t

Table 7: Suffixes of the 8 largest classes

<table>
<tr><th></th><th>PRS</th><th>PST</th><th>IRR</th><th>SUB</th><th>FUT</th><th>IMP</th></tr>
<tr><td>1SG</td><td>1</td><td>10</td><td colspan="3">1</td><td>-</td></tr>
<tr><td>1DU.E</td><td colspan="5">2</td><td>-</td></tr>
<tr><td>1DU.IN</td><td>3</td><td>11</td><td colspan="3">3</td><td>-</td></tr>
<tr><td>1PL.EX</td><td colspan="5">4</td><td>-</td></tr>
<tr><td>1PL.IN</td><td colspan="5">5</td><td>-</td></tr>
<tr><td>2SG</td><td>6</td><td>12</td><td colspan="3">6</td><td></td></tr>
<tr><td>2DU</td><td>7</td><td></td><td colspan="4">7</td></tr>
<tr><td>2PL</td><td colspan="6">8</td></tr>
<tr><td>3SG</td><td>1</td><td>10</td><td colspan="3">1</td><td>-</td></tr>
<tr><td>3DU</td><td>3</td><td>11</td><td colspan="3">3</td><td>-</td></tr>
<tr><td>3PL</td><td colspan="5">9</td><td>-</td></tr>
</table>

Table 8: Suffixal interpredictability areas

4 Discussion

An inflectional system with the complexity of any of these layers would be considered quite complex. The (in)famous Latin verbs, for example, have 4 principal parts, 0.28 bits of average conditional entropies, and 15 zones of interpredictability (see Pellegrini, 2020), yet this is almost consistently simpler than any of the inflectional subsystems that coexist within Pame verbs. The overall system, hence, would appear to test the very limits of human linguistic cognition. How do speakers manage to successfully learn and use a system like this? The answer might lie in predictability *between* inflectional layers. While that between cells is explored more often, as I have done in the previous Section 3, this does not mean that predictability between different slots or properties of a single word plays no role. Preliminary assessment of how much information one layer provides about another in Pame can be obtained from Normalized Mutual Information (NMU), calculated through the R package aricode (Chiquet et al., 2020). Results in Table 9 show NMU oscillates between 0.18 and 0.56, which means the lexical classifications of different layers are highly informative about each other. To mention a few examples, the largest prefixal class is close to incompatible with the absence of stem alternation, the second and third largest prefixal classes are incompatible with tonal alternations, etc.

	tone-stress	stems	suffixes
prefixes	0.261	0.558	0.346
tone-stress		0.270	0.179
stems			0.250

Table 9: NMU between the different slots

Beyond these between-layer predictive relations, another challenging aspect of Pame verb morphology is the unsystematic nature of syncretism. While this is not infrequent in the language (26% of forms), this does not occur systematically, in that there are no cells in the paradigm that are always syncretic. It is remarkable, for example, that prefixal inflection classes (see the largest ones in Table 1) differ not only in their use of different allomorphs, but also in their partition of the semantic space. Because the pattern of contrasts is different in every class of verbs, it must make the Paradigm Cell Finding Problem (see Boyé & Schalchli, 2019) extremely challenging.

A final challenge that Pame verbs present is what Erdmann et al. (2020) have called the Paradigm Identification Problem. Given the amount of suppletion, stem alternation and allomorphy in the system, predicting the lemma and morphosyntactic value of a form from its morphology must also be complicated. The same markers are reused with different functions in different verbs classes (As show in Table 1, for example, la- occurs in the 1SG.PRS prefix in some verbs but as the 2.PRS in other verbs, wa- occurs as a 3SG/DU.PRS in some verbs, but as 3PL.PRS in others, or as 3.PRS or 3.PL in others, etc. These and other aspects can now be explored computationally through the resource VeLePa.

5 Conclusion

This paper has reported on the compilation of a Verbal Lexicon of Pame (VeLePa), specifically with computational applications in mind. It has also presented some preliminary quantitative analyses of this inflectional system around the topics of the Paradigm Cell Filling Problem, and related challenges that speakers and learners of a language face when using and/or acquiring inflectional morphological patterns.

This system, like other Otomanguean ones (see e.g. Cruz et al. 2020) is remarkable because its morphology deviates very significantly and in several dimensions from the canonical (Corbett 2009) most straightforward one. It is structured into several morphological slots which work together (see the phenomena of Multiple, Distributed or Extended Exponence, Harris 2017) into expressing tense-aspect-mood and subject person-number values. Each slot is, furthermore, organized into a large number of inflection classes and contains multiple isolated irregularities.

VeLePa is expected to contribute to both theoretical linguistic analysis and to NLP, allowing the inclusion (e.g. into reinflection tasks) of a language that is both highly complex and typologically very different from the better documented (Indo-)European ones.

Acknowledgments

Thanks are due to the Swiss Society for Endangered Languages and to the University of Zurich's GRC for financial support to conduct the fieldwork on which this research is based. I also thank the patient help of my main informants Aniceto and Josefa. All mistakes are, of course, mine.

References

Ackerman, F., Blevins, J. P. and Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins and J. Blevins (Eds.), Analogy in grammar: Form and acquisition: 54–82. Oxford: OUP.

Beniamine, S. (2018). Classifications flexionnelles. Étude quantitative des structures de paradigmes. PhD diss., Université Sorbonne Paris Cité.

Blevins, J. P., Milin, P. and Ramscar, M. (2017). The Zipfian paradigm cell filling problem. In Perspectives on morphological organization: 139-158. Brill.

Boyé, G., and Schalchli, G. (2019). Realistic data and paradigms: The paradigm cell finding problem. Morphology, 29: 199-248.

Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. Verbal Learning & Verbal Behavior, 10: 722-729.

Chiquet, J., Rigaill, G., Sundqvist, M., Dervieux, V., and Bersani, F. (2020). Package 'aricode'.

Corbett, G. G. (2009). Canonical inflectional classes. In *Selected proceedings of the 6th Décembrettes: Morphology in Bordeaux*:1-11.

Cruz, H., Stump, G., and Anastasopoulos, A. (2020). A resource for studying chatino verbal morphology. *arXiv preprint arXiv:2004.02083*.

Erdmann, A., Elsner, M., Wu, S., Cotterell, R., and Habash N. (2020). The Paradigm Discovery Problem. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*: 7778–7790.

Gibson, L. F. (1950). Verb paradigms in Pame. https://www.sil.org/resources/archives/53023.

Gibson, L. F. (1966). The Man Who Abandoned his Wife. A Pame Story. Tlalocan 5, no. 2: 169-177.

Gibson, L. F., Olson, D. and Olson, A. (1963). Four Pame Texts. Tlalocan 4, no. 2: 125-143.

Harris, A. C. (2017). *Multiple exponence*. Oxford University Press.

Herce, B. (2022). Possessive inflection in Chichimec inalienable nouns: The morphological organization of a closed irregular class. Studies in Lang. 46, 4: 901-933.

Hurch, B. (2022). Pame (central) de Santa María Acapulco, Santa Catarina, San Luis Potosí. In Y. Lastra (Ed.) Archivo de Lenguas Indígenas de México. México DF: El Colegio de México.

McCarthy, A. D., Kirov, C., Grella, M., Nidhi, A., Xia, P., Gorman, K., Vylomova, E. et al. (2020). UniMorph 3.0: Universal Morphology. In Proceedings of The 12th LREC: 3922-3931. ELRA.

Palancar, E. L. and Avelino, H. (2019). Inflectional complexity and verb classes in Chichimec. Amerindia 41: 323-360.

Pellegrini, M. (2020). Using LatInfLexi for an entropy-based assessment of predictability in Latin inflection. In Proceedings of LT4HALA 2020: 37-46.

Stump, G. and Finkel, R. A. (2013). Morphological typology: From word to paradigm. Cambridge: CUP