'Markedness' is an epiphenomenon of random phonetically grounded sound change

Ollie Sayeed, Andrea Ceolin
University of Pennsylvania

1 Introduction

The concept of 'markedness' has been influential in phonology for almost a century, but a recent movement in the field has argued that phonological theories do not need to make reference to the concept; instead, if it is meaningful at all, markedness should be thought of as *emergent*. In this paper, we propose a simple mathematical model based on the principles of Evolutionary Phonology (EP; Blevins, 2004) to explore how a theory without markedness can replicate some of the insights of the markedness-based worldview. We see that markedness can be treated as an epiphenomenon of random, phonetically grounded sound change.

In Section 2, we introduce the various properties that have been claimed to correlate with 'markedness', and the alternative perspective given to us by EP. In Section 3, we introduce the split-and-merger model, an abstract model of sound change based on EP. Section 4-6 describe the predictions of the model in three areas: phoneme frequency distributions (Section 4), correlations between within-language and across-language frequency (Section 5), and contextual distributions (Section 6). We finish by giving a broader picture, in Section 7, of how the different correlates of markedness relate to each other.

2 Background

In the past, theoretical phonology has found it useful to describe some segments as more 'marked' than others, referring to a cluster of language-internal and -external properties (Jakobson, 1941; Haspelmath, 2006):

- Marked segments have low frequency within a language.

- Marked segments have low frequency across languages.

- Marked segments have a more restricted distribution within a language.

- Marked segments are acquired later in infancy.

- Marked segments are more likely to be impaired in cases of aphasia.

The first of these is based on 'textual markedness', where a segment or grammatical construction is marked in some language if it has low frequency in texts. Although Haspelmath's definition strictly only includes *token* frequency, we can extend the concept to *type* frequency: a marked segment should appear in fewer words in the lexicon than an unmarked one. The type-frequency sense of 'markedness' is implicitly assumed in historical linguistics - for example, in arguments that the Proto-Indo-European (PIE) 'plain voiced' stops must have had a marked status on the grounds that they had lower frequency within reconstructed PIE roots - given that, in reconstructed languages like PIE, we only have access to (rough) type frequencies and no token frequencies.

The second is Haspelmath's 'typological markedness', where a segment or construction is marked if it is cross-linguistically rare. For example, /q/ appears in fewer languages than /k/, so under this definition,

/q/ is more marked than /k/. As a more specific case of typological markedness, a marked and an unmarked segment might be related by an implicational universal: if a language allows coda clusters, it allows singleton codas, for example. In these cases, the existence of the marked segment implies the existence of the unmarked one. These first two senses of 'markedness' correspond to within-language and across-language frequencies respectively, and an important empirical observation is that these two kinds of frequency are correlated (Gordon, 2016).

The third is Haspelmath's 'distributional markedness', where a segment or construction is marked if it has a restricted distribution within a language. English /ŋ/, for example, has a narrower phonotactic distribution than the other nasals /m/ and /n/, being banned from onsets.

The last two are from Jakobson (1941), in a discussion of psychological correlates of linguistic markedness. Jakobson claims that children start off acquiring and producing the least marked segments in the input early on in the acquisition period, and only later start producing marked segments. A second claim is that aphasics are more likely to show impairment for marked segments than for unmarked segments. Some recent work (Romani et al., 2017) has shown that both of these are correlated with both within-language and across-language frequencies, as in Haspelmath's senses of 'markedness'.

Starting with Trubetzkoy (1969), there has been a tradition of encoding 'markedness' in the grammar: in the theoretical practice that gives his name to Haspelmath's 'Trubetzkoyan markedness', Trubetzkoy claimed that a 'marked' segment in a pair was literally *marked* (or *specified*) for some phonological distinction, and the unmarked segment was the default. In the decades since, markedness has become central to all mainstream phonological theories, from the universal markedness rules of SPE to the markedness constraints of OT and its descendants.

A recently popular alternative philosophy, EP, has argued that the concept of representational markedness is unnecessary (Blevins, 2004; Hale & Reiss, 2008; Vaux & Samuels, 2017). Instead, EP treats markedness as emergent from diachrony. This corresponds roughly to Haspelmath's 'phonetic markedness', or 'markedness as phonetic difficulty'. As we see in the next section, this phonetic sense of 'markedness' can be encoded in a model of sound change following the assumptions of EP. The broad goal of this paper is to explore how all of the above senses of 'markedness' can be ultimately boiled down to phonetic difficulty.

3 Model: random splits and mergers

In the EP worldview (Blevins, 2004), there are two reasons why segments might be cross-linguistically rare: either sound change tends not to create them very often, or sound change tends to destroy them too often. In other words, either they do not often come into existence, or they do not stay in existence very long when they do. The reason an individual segment has one of these properties could be either articulatory or perceptual - a marked segment is hard to produce or hear (or easy to misproduce or mishear) - and both articulatory and perceptual difficulty are bundled up in Haspelmath's definition of 'markedness' as phonetic difficulty.

We propose a simple abstract model of sound change as a discrete-time stochastic process of random splitting and merging of phonemic categories, roughly following the traditional division of sound changes into splits and mergers. Given a toy 'language' represented by its phoneme inventory, each tick of the clock in our model corresponds to a new sound change acting on the language, which will change the frequency and distribution of some segment in the inventory. The aim is that after each toy language has undergone a sequence of random sound changes, just as natural languages have undergone thousands of years of diachrony, the outcomes should be qualitatively the same as what we see in the world's languages.

Say the phoneme inventory of a language is a set of segments $\{x_i\}$, where the ith segment x_i has frequency p_i^t at time step t. At each stage, we apply either a *split* or a *merger* to the language with equal probability:

- To apply a split, pick a random pair of segments x_i, x_j with $i \neq j$. Take away half of x_i's probability mass, and add it to the existing probability mass of x_j. Leave all other x_k with $k \neq i \neq j$ alone.

$$p_i^{t+1} := \frac{p_i^t}{2}$$

$$p_j^{t+1} := \frac{p_i^t}{2} + p_j^t$$

$$p_k^{t+1} := p_k^t$$

- Mergers follow a similar algorithm, except that *all* of x_i's probability mass is transferred to x_j.

$$p_i^{t+1} := 0$$

$$p_j^{t+1} := p_i^t + p_j^t$$

$$p_k^{t+1} := p_k^t$$

This corresponds more or less to the traditional concepts of 'split' and 'merger' in historical linguistics. In our model, a split literally splits the probability mass of a segment in two, transferring half of it to another segment, which could be a segment already existing in the language or a new segment with a previous frequency of zero (therefore, following Hoenigswald (1965), a 'primary' or a 'secondary' split). A merger takes two segments and merges them into one, with a new, combined probability mass under the label of one of its parents.

What about markedness? In EP, 'marked' segments are those that have either a low probability of being created by sound change or a high probability of being lost by sound change. To encode these in our model, we can modify the splitting and merging algorithm to include two quantites we call 'splitwise markedness' and 'mergerwise markedness':

- Define a function $P_S(x_j)$ such that $P_S(x_j) \geq 0$ and $\Sigma P_S(x_j) = 1$; this is a probability distribution representing the probability that the jth segment x_j will be created when another segment splits. When the splitting algorithm calls for picking a random pair of segments x_i, x_j, pick x_j randomly according to the distribution $P_S(x_j)$.

- Define a second probability distribution $P_M(x_i)$, representing the probability that x_i is lost in a merger. When the merging algorithm calls for picking a random pair of segments x_i, x_j, pick x_i randomly according to $P_M(x_i)$.

Say that segments with *low* $P_S(x_j)$ are 'splitwise marked', and segments with *high* $P_M(x_i)$ are 'mergerwise marked'. In other words, marked segments *do not* tend to be created in a split (hence the low number), but *do* tend to be lost in a merger (hence the high number). These two senses of markedness in our model are a way of formalizing the concepts from EP.

4   Predictions: phoneme frequency distributions

Before we get into the details of deriving correlates of markedness, one interesting qualitative prediction made by the split-and-merger model is about the *shape* of the frequency distribution for segments within a language. A language that made maximal use of the information capacity of its channel would use every segment with equal frequency; but natural languages are redundant, with some segments used much more frequently than others. In the case of word frequencies, the plot of frequency against rank famously follows a Zipf distribution (Zipf, 1935, 1939), where the rth most frequent word in a language has a frequency proportional to some negative power of its rank:

$$f_Z(r) = \frac{a}{r^b}$$

Empirically, b is close to one, so the frequency of the rth most frequent word is roughly proportional to $1/r$; a is just a normalizing constant to make the frequencies sum to one.

For phonemes, Tambovtsev and Martindale's (2007) survey of corpora from 95 languages finds that a slightly different curve is a better fit for phoneme token frequencies:

$$f_{YS}(r) = \frac{a}{r^b} c^r$$

This is the Yule-Simon distribution (Yule, 1925), with a new parameter $c \leq 1$, which gives us the Zipf distribution as a special case when $c = 1$. We can think of the Zipf distribution as an approximation

to the more general Yule-Simon for a big enough set of objects. As r gets larger, the difference $\frac{a}{r^b}(1 - c^r)$ gets smaller, so the tail of the Yule-Simon distribution is close to the tail of the Zipf distribution. For lexicons containing hundreds of thousands of words, the Zipf plot is a reasonable fit; for phoneme inventories containing a few dozen phonemes, the Yule-Simon plot is more accurate.

Although Tambovtsev & Martindale (2007) only show that *token* frequencies follow a power-law distribution, Martin (2007) finds qualitatively the same rules for *type* frequencies at three stages in the history of English. The difference is important, because our model of sound change only makes predictions about the latter; nothing in our proposal determines which words should be used more or less frequently than other words.

The split-and-merger model is a variant on the class of 'random fragmentation and aggregation processes', which Banavar et al. (2004) show have power-law frequency distributions over the elements being split and merged as their fixed points. In other words, randomly splitting and merging categories arranged in a power-law distribution will give back the same power-law distribution. What is more, a random fragmentation and aggregation process starting with a flat distribution will converge on a power-law distribution and stay there.

Our model is importantly different in involving 're-merging', where a category that splits in half will transfer half of its frequency mass somewhere else, while Banavar et al.'s process leaves each of the two halves as its own smaller category. As we previously said, in terms of Hoenigswald's (1965) classification of sound changes, our process models conditioned mergers ('primary splits'), where the segment produced by a split already exists in the language; a process without re-merging can only model phonemic splits ('secondary splits'), producing a new segment. Our model handles secondary splits by treating the creation of a new segment as the transfer of frequency mass to a category with frequency zero, so the split-and-merger model is more general than the random fragmentation and aggregation process.

We ran a simulation of the split-merger process for 500 iterations with an arbitrary set of 20 phonemes arbitrarily labeled $\{a, b, c,....t\}$ as the starting inventory. We assume that phoneme frequencies are uniform as a starting point: each phoneme has a starting within-language frequency of 0.05.[1] In addition, six phonemes $\{u, v,...z\}$ are assigned an initial value of zero, which means that they are not present in the starting inventory, but can eventually emerge during the iterations as a result of splits. In the sound change simulation, either a split or a merger is applied to the phonemic inventory at each iteration with probability 0.5 each. As previously said, mergers combine together the frequencies of two phonemes chosen at random in the set $\{a, b, c, ...z\}$ into a single phoneme, while splits target one phoneme at random in the same set and move half of its frequency index into another phoneme randomly selected. The algorithm is implemented using Python.

Simulations of the split-and-merger model in action show long-tailed distributions emerging out of an initial flat distribution, qualitatively in line with the results from random fragmentation and aggregation (Figure 1).

If random splitting and merging of phonemic categories predicts power-law frequency distributions, this suggests a diachronic explanation for a synchronic pattern: phoneme frequency distributions in natural languages are a product of random sound change.

5   Predictions: within-language and across-language frequency

As well as predicting the shape of a synchronic frequency distribution, the split-and-merger model predicts a correlation between the first two of our correlates of markedness in Section 1 - frequency within a language (type or token), and frequency across languages. It has long been informally assumed that crosslinguistically rare segments should also show a relatively low frequency in the languages that have them, and more recent work seems to back this correlation up empirically. Gordon (2016) surveys synchronic phoneme frequency data in 32 languages taken from the World Atlas of Language Structures (WALS), and finds that within-language and across-language frequencies correlate well with each other.

The correlation is not perfect, which Gordon puts down to a handful of reasons, including linguists' arbitrary choices in how to transcribe an ambiguous segment - but an interesting group of exceptions have a diachronic explanation. Obviously, within-language frequency will tend to be lower than across-language frequency: within-language frequencies sum to one, while across-language frequencies do not. If a language

---

[1]   We also tried to impose other distribution, like the Zipfian, without any change in the results.
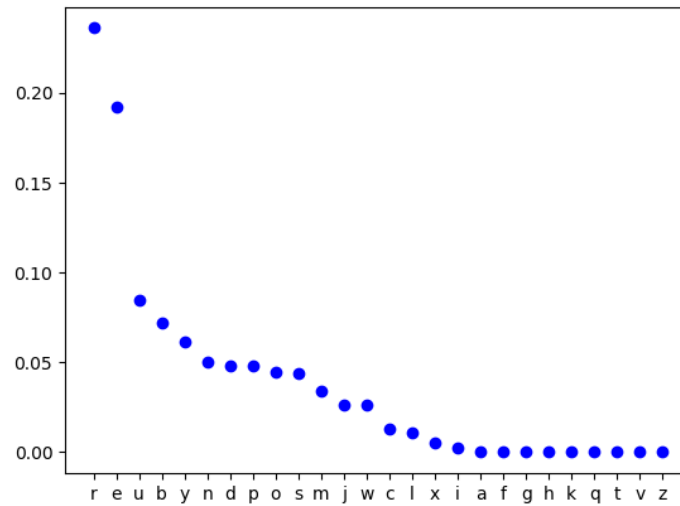
Figure 1. A typical run of our simulation after 500 iterations.

completely lacks /x/ - including when it loses /x/ by sound change - then its within-language frequency of zero will be lower than what we would predict based on its crosslinguistic frequency. In cases where a segment has a *higher* frequency within a language than we would predict, we also have sound change as a potential cause: Gordon gives the example of Romanian, which has /r/ at an unusually high frequency thanks to a recent merger of intervocalic /l/ with inherited /r/. In our model, individual exceptions have individual diachronic explanations, but the *average* result of generations of sound change is an observable correlation between the two kinds of frequency.

In EP, if the markedness of a segment is an epiphenomenon of its diachronic behaviour, we have four facts to explain:

1. Why mergerwise marked segments are rare across languages

2. Why splitwise marked segments are rare across languages

3. Why mergerwise marked segments have low frequency within a language

4. Why splitwise marked segments have low frequency within a language

These fall out as predictions of our model. In other words, the correlation between across-language and within-language frequency appears diachronically because both are consequences of a third variable: 'markedness' under the phonetic biases involved in sound change.

The first two of these facts are easy to see. Suppose that we run the model lots of times to get a representative typology of languages, each time starting with an inventory consisting of a uniformly random sample of segments. If /x/ has a higher probability of being lost through a merger than /y/, we predict that /x/ should survive in fewer runs of the model than /y/, *ceteris paribus*; and since the probability of being lost corresponds to the mergerwise markedness $P_M(x_i)$, we see that more mergerwise marked segments are predicted to be rarer.

In a mirror image of this reasoning, if /x/ has a *lower* probability of being *created* by a split than /y/, we predict /x/ should appear in fewer runs than /y/, *ceteris paribus*; and since the probability of being created corresponds to $P_S(x_i)$, we can see that more splitwise marked segments (with low $P_S(x_i)$ are also predicted to be rarer.

5

And so both types of markedness predict crosslinguistic rarity. To show that marked segments should be rarer *within* a language, we need to use some probability theory - the claim is that marked segments should rarer than unmarked segments within an *average* run of the model.

We can see this by writing down an expression for the expected value $E(p_i^{t+1})$ of the ith segment at time $t + 1$ as a function of its frequency at time t and the dynamics of the model:

$$E(p_i^{t+1}) = \frac{1}{2}\left(P_M(x_i)\frac{1}{N}p_i^t + P_M(x_i)\frac{N-1}{N}\left(p_i^t + \sum_{j\neq i}\frac{1}{N-1}p_j^t\right) + (1 - P_M(x_i))\frac{N-1}{N}p_i^t\right)$$

$$+\frac{1}{2}\left(\frac{1}{N}P_S(x_i)p_i^t + \frac{1}{N}(1 - P_S(x_i))\left(p_i^t + \sum_{j\neq i}P_S(x_j)\frac{p_j^t}{2}\right) + \frac{N-1}{N}P_S(x_i)\frac{p_i^t}{2} + \frac{N-1}{N}(1 - P_S(x_i))p_i^t\right)$$

At each time step, with probability $\frac{1}{2}$, we either perform a random split or a random merger, each of which involves two out of N segments. The terms in this expression correspond to possible choices of these two segments and their effects on the probability of the ith segment on the next time step. Suppose a merger happens, for example:

- With probability $P_M(x_i)$, we pick x_i as the segment that survives the merger, and with probability $\frac{1}{N}$, x_i is also the segment 'lost' in the merger. In this scenario, x_i is 'merging' with itself and nothing changes, so its frequency at the next time step will stay the same at p_i^t.

- If instead some other segment is merged with x_i, which happens with probability $\frac{N-1}{N}$, then the frequency of x_i will increase by whatever the frequency of the other segment was. The average frequency by which x_i will increase, summed over all $N - 1$ other segments, is $\sum_{j\neq i}\frac{1}{N-1}p_j^t$; so the expected frequency of x_i after the merger is $p_i^t + \sum_{j\neq i}\frac{1}{N-1}p_j^t$.

- If x_i isn't picked as the surviving segment, but *is* picked as the segment to be lost, its frequency in the next generation will be zero; so the corresponding term disappears from the sum.

- Finally, if the merger doesn't involve x_i at all, which happens with a probability of $(1 - P_M(x_i))\frac{N-1}{N}$, x_i isn't affected, and its frequency will stay at p_i^t.

The expected frequency of x_i after a merger, then, is just a sum of each possible outcome weighted by its probability. What we want to do is see how this varies as a function of the mergerwise markedness $P_M(x_i)$. Some simple algebra shows that, by rearranging terms in our expression for the expected frequency, the coefficient of $P_M(x_i)$ is $\frac{1}{2N}$. This quantity is always positive, and so $E(p_i^{t+1})$ goes up when $P_M(x_i)$ goes up. In other words, segments with low mergerwise markedness are more frequent in the average language than segments with high mergerwise markedness. An analogous line of reasoning based on the second term of this expected frequency shows that the same is true for splitwise markedness: low splitwise markedness means higher frequency.

In order to visualize this result, we ran another simulation of the split-merge process for 500 iterations, implementing mergerwise markedness. This time, /a/ is 'umarked' with respect to the other segments, and /b/ is 'marked'; so mergers are biased towards preserving /a/ and deleting /b/. In a typical run, as expected, /a/ becomes high frequency and /b/ falls out of the language (Figure 2).[2]

We can plot the average frequencies of /a/ and /b/ across 1000 parallel runs given equal starting frequencies of 0.05; /a/ has an average frequency of 0.11, while /b/ disappears in most runs (Figure 3). This behaviour is statistical, not absolute: the figure shows that there are cases in which /a/ disappears and cases in which /b/ ends up higher than the starting frequency, despite evolutionary biases. This follows from the fact that the two phonemes are the same in terms of splitwise markedness in this first experiment: it means that /a/ can lose probability mass because of a split, and /b/ can gain probability mass as a consequence of splits affecting other segments.

---

[2] In the implementation, the likelihood of being loss after a merger is zero for /a/, and uniform for all the segments but /b/, whose probability is double than that of the other segments.
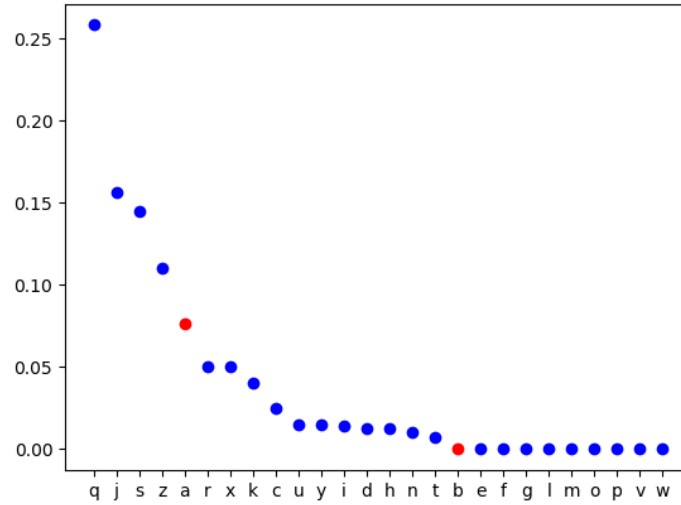
Figure 2. A typical run of our simulation after 500 iterations. /a/ is mergerwise unmarked, while /b/ is mergerwise marked.
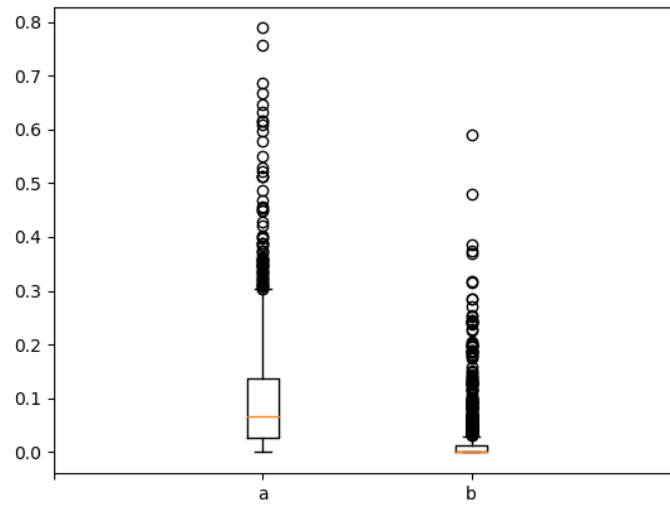
Figure 3. Within-language frequencies of /a/ and /b/ across 1000 parallel runs.

Phoneme	Mergerwise Markedness Bias	Within-language frequency	Across-language frequency
/a/	Unmarked	0.109	1.0
/c/	-	0.071	0.51
/b/	Marked	0.059	0.331

Table 1. Average within- and across-language frequencies for three phonemes which differ in terms of markedness.

At this point, not only can we track the average frequencies of the two phonemes in the languages in which they are present (within-language frequencies), but also keep track of the number of languages in which they survive (across-language frequencies), and compare them with the frequencies exhibited by phonemes which are neither 'unmarked' nor 'marked', for example /c/. Importantly, within-language frequencies are calculated only for languages in which the segment was present at the end of the run.

Table 1 shows correlation between within- and across-language frequencies. First, /a/ had an average frequency of 0.109 in the runs in which it survived, namely all of them. Given the way our model works, the bias towards /a/ makes it impossible for it to disappear from the language as a result of a merger, and frequencies lower than the initial one (0.05) can only be achieved by means of splits, which always remove half of the probability mass, and therefore they make it impossible to reach zero for the phoneme that they target. However, /c/, which did not have any bias in terms of mergers, had an average frequency of 0.071 and it disappeared in about half of the runs. Finally, /b/, which was mergerwise 'marked', showed an average frequency of 0.059, and it disappeared in about two thirds of the runs.

To sum up, these simulations give us some qualitative corroboration of what we knew anyway: when we add a diachronic bias to the split-and-merger model, we see within-language and across-language frequency lining up.

6   Predictions: synchronic distributions

A more speculative application of the model comes in explaining Haspelmath's 'distributional' markedness, where a segment is marked within a language if it has a restricted distribution within words. If a segment only appears in onsets, for example, it's more distributionally marked than a segment that can appear in all positions.

This property cannot be encoded in the model we have presented so far, because the split-and-merger model applies sound changes to phoneme inventories - not to sequences of phonemes. To handle the effect of sound change on the synchronic distributions of these phonemes in the language, we would need to make two adaptations to the model. One is that, rather than just a list of phonemes and their frequencies, each state of the language will also need to encode a list of phonotactic statements about the allowable distribution of each phoneme: we could do this by applying each sound change to a lexicon of strings, and saying a sequence is phonotactically allowable if it appears somewhere in the lexicon. The second is that sound changes need to have some conditioning environment; otherwise, all instances of a segment will change in the same way under an unconditioned sound change, and so we would never see a segment appearing in some environments but not others.

Intuitively, the model should predict that both types of markedness will correlate with a more restricted distribution in the average language. If each split is conditioned by some environment, then a segment that rarely tends to be created by a split will also tend to appear in fewer environments; and if each merger is conditioned by some environment, then a segment that often tends to be lost through a merger will also tend to lose parts of its distribution over time. A more rigorous argument, though, will depend on having a more specific description of how we encode phonotactics and conditioned changes into the split-and-merger model.

7   Other correlates of markedness

What about the other two correlates of markedness: age of accquisition, and probability of impairment in cases of aphasia? Romani et al.'s (2017) study of Italian finds that these two measures are both correlated with
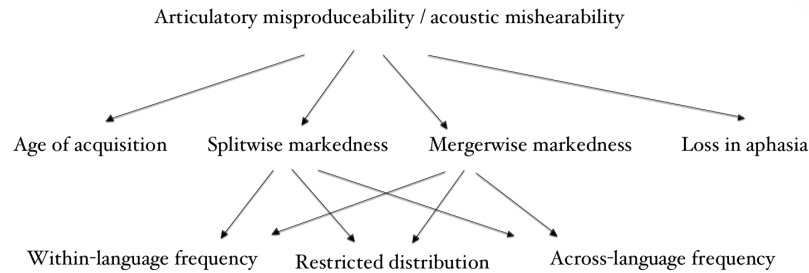
Figure 4. Summary graph.

both types of frequency we saw in section 4. Interestingly, the correlation with cross-linguistic frequency is *independent* of the correlation with language-internal frequency; so we can't reduce this to the statement that children tend to acquire frequent segments earlier, or that aphasics tend to lose infrequent segments.

Instead, age of acquisition and probability of impairment have to be correlated with the two types of frequency via some other factor. Romani et al. treat this as 'articulatory difficulty', to which we can add perceptual difficulty - of the type that also underlies biases in sound change. In our model, following EP, 'markedness' really reduces to these two types of phonetic difficulty, which in turn will have explanations grounded in the physics of the mouth and the psychology of perception.

So the causal relationships between correlates of markedness start off with phonetic factors, and ultimately trickle downstream into synchronic tendencies (Figure 4).

8   Conclusion

As we have seen, both the power-law frequency distribution of phonemes in a language and the cluster of properties associated with 'markedness' can be thought of as epiphenomena of diachrony. A simple model of sound change consisting of random phonetically biased splits and mergers does the same job for us in predicting the attested language-internal and typological correlations.

References

Banavar, Jayanth R, Paolo De Los Rios, Alessandro Flammini, Neal S Holter & Amos Maritan (2004). Scale-free behavior and universality in random fragmentation and aggregation. *Physical Review E* 69:3.

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Gordon, Matthew K (2016). *Phonological typology*. Oxford University Press.

Hale, Mark & Charles Reiss (2008). *The phonological enterprise*. Oxford University Press.

Haspelmath, Martin (2006). Against markedness (and what to replace it with). *Journal of linguistics* 42:1, 25–70.

Hoenigswald, Henry M (1965). *Language change and linguistic reconstruction*. University of Chicago Press.

Jakobson, Roman (1941). *Child language, aphasia and phonological universals*. Walter de Gruyter.

Martin, Andrew Thomas (2007). *The evolving lexicon*. Ph.D. thesis, University of California.

Romani, Cristina, Claudia Galuzzi, Cecilia Guariglia & Jeremy Goslin (2017). Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive neuropsychology* 34:7-8, 449–471.

Tambovtsev, Yuri & Colin Martindale (2007). Phoneme frequencies follow a Yule distribution. *SKASE, Journal of Theoretical Linguistics* 4:2, 1–11.

Trubetzkoy, Nikolai S (1969). *Principles of phonology*. University of California Press.

Vaux, Bert & Bridget D Samuels (2017). Consonant epenthesis and markedness. Samuels, Bridget D (ed.), *Beyond Markedness in Formal Phonology*, John Benjamins Publishing Company, 69–100.

Yule, George Udny (1925). A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London Series B* 213, 21–87.

Zipf, George Kingsley (1935). *The psycho-biology of language: An introduction to dynamic philology*. Routledge.

Zipf, George Kingsley (1939). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley, Cambridge, MA.