Are Universal Markedness Hierarchies Learnable from the Lexicon? The Case of Gemination in Hungarian

Lilla Magyar

1. Introduction

Gemination in loanwords is a cross-linguistically widespread phenomenon: a singleton consonant in the source word is geminated in the loanword, even if the doubling does not have an orthographic reflex (that is, the geminated consonant is spelt with a single consonant letter in the source word). Source languages which these words are borrowed from do not allow phonetic geminates. Some examples of gemination in loanwords are listed below:

- Japanese: [kat:o] `cut' (Kubozono et al. (2008))
- Italian: [fan:] `fan' (Passino (2004))
- Kannada: [kap:u] `cup' (Sridhar (1990))
- Telugu: [ro:ɖːu] `road' (Krishnamurti & Gwynn (1985))
- Finnish: [pop:i] `pop' (Karvonen (2009))
- Hungarian: [ʃokː] `shock' (Nádasdy (1989), Kertész (2006))

In Japanese, Kannada, Telugu and Finnish, the final consonant of the source word is geminated, and a vowel is inserted after the final consonant. Loanword gemination in Italian and Hungarian is slightly different from what we see in other languages. In Italian, the final consonant of monosyllabic words is geminated, which is very similar to the Hungarian loan gemination process.

The mechanism of loanword gemination oftentimes differs from gemination processes in the native phonology. In Hungarian, for example, all consonants can be geminated in the native phonology, but gemination in loanwords is much more restricted: only certain consonants can undergo loan gemination. However, it is not clear whether gemination in the native phonology is also more restricted than it seems to be at first sight, that is, whether certain consonants are only geminated across morpheme boundaries or in certain positions in the word. The goal of this paper is to explore the differences and similarities between native and loan gemination processes; how they are related to the question of cross-linguistic geminate markedness; whether native speakers of Hungarian have awareness of universal geminate markedness, and if so, is it possible that they draw their generalisations from the native phonology.

1.1. Gemination

In Hungarian, gemination is possible in the native phonology as well as in loanword adaptation. This section is a review of these two types of gemination, and their relation to cross-linguistic patterns of geminate markedness.

[*] Lilla Magyar, MIT, lilla@mit.edu. I am grateful to Adam Albright, Donca Steriade, Michael Kenstowicz, Edward Flemming, Miklós Törkenczy, Katalin Mády, Csaba Oravecz, and audiences at the 22nd Manchester Phonology Meeting, the Phonology Circle at MIT and the 33rd West Coast Conference on Formal Linguistics for valuable comments and discussion. All remaining errors and inaccuracies are attributed to me alone.

© 2016 Lilla Magyar. Proceedings of the 33rd West Coast Conference on Formal Linguistics, ed. Kyeong-min Kim et al., 287-298. Somerville, MA: Cascadilla Proceedings Project.

1.1.1. Gemination in loanwords

In Hungarian, there are different contexts for gemination in loanwords, but it is most common in word-final position in monosyllabic words. In `recent' (from 1750 onwards) Hungarian loanwords borrowed from English, German (and occasionally, from French), short consonants following a short (usually stressed) vowel in the source word are regularly geminated in the loanword, even if the source word does not contain a double consonant letter (Nádasdy (1989), Kertész (2006)). Loan gemination without orthographic reflex in the source word is most common in word-final position in monosyllables.

Apart from position in the word, gemination in loanwords depends on consonant class as well. For example, voiceless obstruents are geminated more often than other consonants; nasals are geminated more often than liquids; and voiced fricatives are never lengthened unless the source word contains a double consonant letter in the spelling.

1.1.2. Gemination in the native phonology

The Hungarian consonant inventory consists of 25 consonants. All consonants can be geminated, which indicates that, unlike in the case of loanwords, there seem to be no restrictions on gemination by consonant class in the native phonology. However, some consonants may be geminated only across morpheme boundaries. The frequency distribution of singletons and geminates in the native phonology is not clear. Geminates can occur in both intervocalic and word-final position. Furthermore, gemination is contrastive: it is possible to find minimal or near minimal pairs which are differentiated by consonant length.

1.1.3. Gemination cross-linguistically

The typology of geminate markedness by consonant type has been widely discussed in the literature (Thurgood (1993), Morén (1999), Podesva (2002), Kawahara (2007), Kraehenmann (2011)), but findings are contradictory: some studies report that there is preference for sonorant geminates (Kawahara (2007)), whereas others claim that there are no universal preferences for gemination by consonant type (Morén (1999)). Although there is no complete markedness hierarchy of geminates across languages, some partial rankings and implicational hierarchies have been reported by studies (Podesva (2002), Steriade (2004)). In particular, the following asymmetries appear to hold cross-linguistically: If a language allows geminates, it is most likely to have geminate voiceless obstruents. In a language which permits geminate sonorants, geminate nasals are more likely than geminate liquids, and geminate liquids are more common than geminate glides. Geminate voiced fricatives are very rare cross-linguistically.

The cross-linguistic picture is very similar to what we can observe in the Hungarian loan gemination process (Nádasdy (1989)). Voiceless obstruents are the most frequently geminated consonants in loanwords. Nasals are more often geminated than liquids. Voiced fricatives are never geminated unless the source word contains a double consonant letter.

1.2. Questions and hypotheses

Since gemination in loanwords seems to reflect preferences of cross-linguistic geminate markedness, the question arises whether native speakers of Hungarian have awareness of such hierarchies. However, we have shown that all consonants can be geminated in the native phonology. If that is the case, where does speakers' knowledge of geminate markedness come from? If we take a closer look at the distribution of Hungarian geminates and singletons, would we find evidence that such asymmetries play a role in the native phonology, too? If the frequency distribution of geminates by consonant class in the native phonology does indeed line up with patterns of universal geminate markedness and native speakers are aware of these generalisations, is it possible that these preferences are learnt from the native lexicon, or do we need to posit that speakers have additional knowledge of universal markedness asymmetries? Based on the above discussion, we will formulate and test the following hypotheses:

- Hypothesis 1: Native speakers of Hungarian (a language which allows all kinds of geminates) have awareness of universal geminate markedness.
- Hypothesis 2: This knowledge mirrors the native lexicon: the frequency distribution of geminates in the native phonology reflects patterns of universal markedness.
- Hypothesis 3: These patterns can be learnt from the native Hungarian lexicon based on phonotactic generalisations.

2. Testing Hypothesis 1: Do native Hungarian speakers have awareness of universal geminate markedness?

In this section, we are going to test Hypothesis 1, that is, whether native speakers of Hungarian - a language which seemingly has no restrictions on gemination by consonant class - have some sort of awareness of cross-linguistic geminate markedness.

2.1. Wug test

We conducted a nonce well-formedness judgement test in order to find out whether native speakers of Hungarian prefer certain geminates or singletons, and whether their preferences line up with asymmetries of loanword gemination and cross-linguistic geminate markedness.

2.1.1. Experiment

We used monosyllabic nonce words as test items.[1] The test contained 236 target items (118 word pairs): all the possible combinations of short vowels and short consonants or geminates. All words ended in a short vowel and either short consonant or geminate sequence. In addition to these, filler items - monosyllables containing a long vowel followed by a short consonant - were also included.

115 native speakers of Hungarian participated in the experiment, who were recruited on Facebook and various university mailing lists. They volunteered for the experiment and were not paid. All of them live in Hungary and none of them have spent more than two years in an English- or a Geman-speaking country.

The task was administered online and participants remained anonymous. Participants were presented with a word pair and asked to decide which member of the word pair sounds more plausible as a Hungarian word or a Hungarianised loanword. All word pairs were minimal pairs, containing a monosyllable ending in a short vowel + singleton sequence and another one with the same vowel and the same consonant in a geminated form (e.g. *mok* [mok] - *mokk* [mok:]). The order of presentation was balanced, that is, sometimes the word with the singleton came first, sometimes the one with the geminate. Although the stimuli were written, not spoken, participants were asked to treat spelling as a strict representation of pronunciation.[2]

[1] Gemination in loanwords borrowed into Hungarian is most common and almost only productive in monosyllabic words, following short vowels.

[2] Generally, when native speakers of Hungarian see a double consonant letter in spelling, they tend to pronounce it as a geminate, unless there are some degemination rules at play due to contextual restrictions (e.g. word-final double consonant preceded by a long vowel, geminate surrounded by other consonants).

2.1.2. Results

Table 1 illustrates the results of the wug test.

Consonant class	Geminate %
Voiceless affricates	**56%**
Voiceless stops	**53%**
Voiceless fricatives	**53%**
Liquids	**52%**
Nasals	**39%**
Voiced stops	**35%**
Voiced fricatives	**15%**

Table 1: Wug test results

The first column contains the consonant classes, whereas the second one shows the percentage of responses preferring geminates. The results indicate that native speakers of Hungarian generally prefer voiceless obstruent geminates to other long consonants, and voiced fricatives are the least preferred geminates. However, geminate liquids are exceptionally highly rated.[3] Apart from the preference for geminate liquids, native speakers' judgements seem to line up with patterns of cross-linguistic geminate markedness and gemination in loanwords borrowed into Hungarian, that is, preference for voiceless obstruent geminates over other kinds of geminates and clear dispreference for voiced fricatives.

3. Testing Hypothesis 2: Is cross-linguistic geminate markedness reflected in the native lexicon?

Results of the previous section show that speakers have `active' (generalisable) knowledge that some segments are more likely to be geminated than others. Where does this knowledge come from? Does it reflect knowledge of universal markedness hierarchies? Or could it be learnt from Hungarian-internal data somehow? In this section, we are testing Hypothesis 2, that is, whether patterns of cross-linguistic geminate markedness (and loanword gemination) are reflected in the native lexicon.

3.1. Corpus study

We extracted all monosyllables containing a short vowel followed by a consonant (singleton or geminate) from the Hungarian Webcorpus (Halácsy et al. (2004)). Very recent loanwords (which are considered to be `foreign' by native speakers) were excluded. We counted the type frequency of words ending in singletons and geminates, and collapsed the results by consonant class. This is shown in Table 2.

Consonant class	Geminate %
Voiceless affricates	**65%**
Voiceless stops	**49%**
Voiceless fricatives	**46%**
Voiced stops	**35%**
Nasals	**22%**
Liquids	**28%**
Voiced fricatives	**8%**

Table 2: Corpus counts

[3] Participants in the experiment reported that many items containing geminate liquids looked like famous brand names or endings of common polysyllabic words.

The first column contains consonant classes, while the second column shows the distribution of geminates in the corpus. We can see that voiceless obstruents are the most common geminates. Geminate nasals are somewhat more common than liquids. Geminate voiced fricatives are extremely rare in the native Hungarian lexicon.

3.2. Comparison with wug test results

There is a strong positive correlation (r=0.85) between native speakers' judgements (wug test results) and type frequency counts in the corpus, which is plotted as Figure 1.

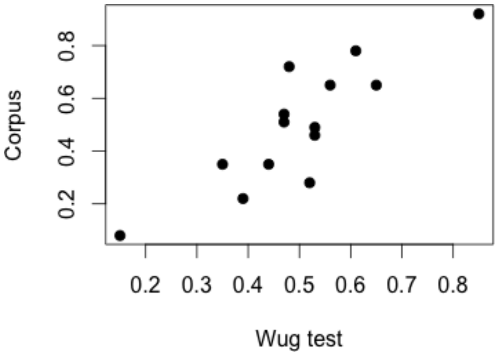

Figure 1: Correlation between wug test results and corpus counts

Both native speakers' judgements and corpus counts reflect universal tendencies in some ways. In each case, voiceless obstruent geminates are more common than other geminates. Geminate voiced fricatives are hardly ever preferred over singletons by native speakers, and they are extremely rare in native Hungarian phonology as well. However, there are some differences. Nasal geminates are more frequent than liquid geminates in the corpus (as observed cross-linguistically), but liquid geminates are unusually highly rated in native speakers' judgements.

4. Testing Hypothesis 3: Can cross-linguistically observed patterns of markedness be learnt from the native lexicon?

We have seen in the previous sections that both native speakers' preferences and type frequencies in the corpus line up with patterns of cross-linguistic geminate markedness and gemination in loanwords borrowed into Hungarian. Since Hypothesis 1 and Hypothesis 2 have been confirmed, the question arises whether Hypothesis 3 holds, that is, whether these preferences (and markedness constraints) can be learnt from the lexicon based on phontactic generalisations. In this section, we are going to test Hypothesis 3.

4.1. The model

The model is implemented using the UCLA Phonotactic Learner (Hayes & Wilson (2008)). It learns and weights constraints. The weighting of constraints is based on the principle of Maximum Entropy.[4]

4.1.1. How it works

Given a set of data (e.g. words extracted from a corpus), the learner creates markedness constraints based on phonotactic generalisations. The probability of each form (e.g. word) is assigned based on the sum of the violations for all of the constraints. These sums - harmony scores - can be converted into probabilities in two ways:

[4] Maxent models are log-linear models widely used in many fields including natural language processing and theoretical linguistics (see Berger et al. (1996), Rosenfeld (1996), Della Pietra et al. (1997), Jelinek (1999), Manning & Schütze (1999), Eisner (2000), Eisner (2001), Klein & Manning (2003), Goldwater & Johnson (2003), Jäger (2004) and Hayes & Wilson (2008), amongst many others).

1. If we want to calculate the probability of a form (e.g. *ibb* [ib:]) given all the other forms, we have to look at global probability, which is calculated as follows:

$$P(ibb) = \frac{exp(-h(ibb))}{exp(-h(ibb)) + exp(-h(opp)) + exp(-h(et)) + ...}$$

$$P(ibb) + P(ib) + P(opp) + P(et) + ... = 1$$

2. The probability of a form given another form (e.g. the probability of *ibb* [ib:] given *ib* [ib]), that is, local probability, is calculated as follows:

$$P(ibb) = \frac{exp(-h(ib))}{exp(-h(ibb)) + exp(-h(ib))}$$

$$P(ibb) + P(ib) = 1$$

Local probability provides a better comparison with the wug data, because a direct comparison of geminate and singleton forms corresponds more closely to the task participants were asked to perform in the experiment. Therefore, we will convert harmony scores to local probabilities.

4.2. Simulation 1

The goal of Simulation 1 was to see whether geminate markedness is learnable when the learner is provided with full phonetic detail.

4.2.1. Settings

Monosyllabic lemmas were extracted from the Hungarian Webcorpus (Halácsy et al. (2004)) and used as training data in Simulation 1. Rhymes (short vowel + consonant sequences) from the wug test were used as testing items. Apart from these data, a feature chart was given to the learner, which contained distinctive features characterising Hungarian consonants and vowels. The simulations were run both with and without a fixed number of constraints. The results reported are representative of several runs.

4.2.2. Results

The probabilities assigned by the learner were compared to wug test judgements to see how much a word ending in a certain short vowel + short consonant sequence is (dis)preferred over a sequence (or a full word) ending in the same vowel + geminate sequence (e.g. *op* [op] vs. *opp* [op:]).

In Figure 2, predicted probabilities (scores assigned by the model converted to local probabilities) are shown on the *x*-axis, while observed probabilities (wug test results) are plotted on the *y*-axis. There is a very weak positive correlation (r=0.25) between the two results, which means that native speakers' judgements on various monosyllables do not line up with type frequencies of their VC rhymes in the corpus. This is illustrated by Figure 2.

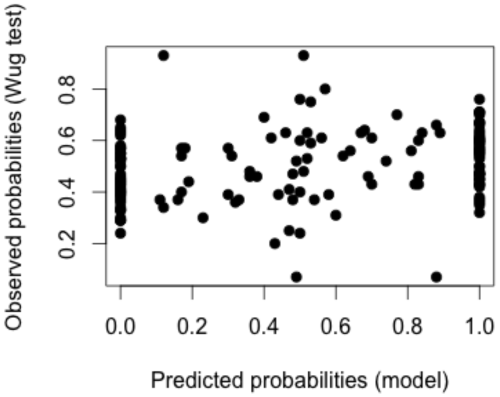

Figure 2: Simulation 1

The reason for this could also be that the learner failed to make correct generalisations from the corpus. However, this is not the case. If we compare actual type frequency distributions with probabilities assigned by the learner, we can see that the correlation is very high: a perfect or an almost perfect match.

In Simulation 1, the learner had to focus on several details like consonant-vowel co-occurrences (both onset-nucleus and nucleus-coda) and restrictions unrelated to the question of gemination, which might also be accidental gaps. Therefore, the learner might not have been able to focus on the distribution of geminates and singletons. Moreover, this might not be the problem of the learner but something specific to the process of how native speakers generalise: they might not take the whole word form into consideration in judging whether they prefer geminated or singleton forms, but they concentrate on rhymes instead.

4.3. Simulation 2

Based on generalisations drawn from Simulation 1, the learner will be trained on rhymes instead of full lemmas in this simulation.

4.3.1. Settings

In Simulation 2, the learner was trained on all rhymes (vowel + final consonant sequences) of monosyllables extracted from the corpus and tested on final consonants (singletons and geminates) from the wug test.

4.3.2. Results

Predicted probabilities of final singletons and geminates were compared to wug test judgements (that is, how much words ending in a certain singleton (e.g. [b]) are preferred or dispreferred over a geminate (e.g. [b:]). The correlation between the learner's predictions and native speakers' preferences (r=0.14) is even weaker than in Simulation 1. This is plotted in Figure 3.

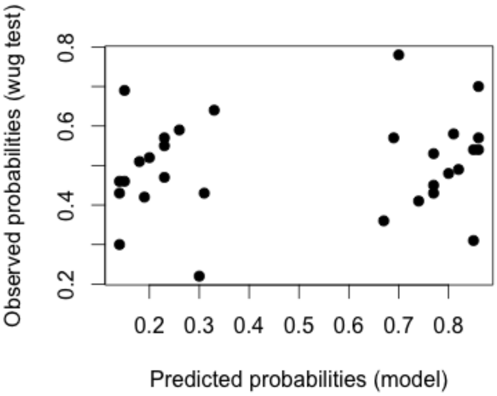

Figure 3: Simulation 2

It is clearly made visible by Figure 3 that native speakers' judgements (averaged by final VC sequences of monosyllables) do not match up with the distribution of rhymes in the corpus of monosyllables.

4.4. Simulation 3

Following the same logic as above, one possible interpretation of the failure of Simulation 2 is that fewer pieces of information are taken into consideration in the learning process. Therefore, only final consonants will be taken into consideration in this simulation.

4.4.1. Settings

In Simulation 3, the learner was trained on word-final consonants extracted from the corpus and tested on word-final consonants in the wug test. The learner was also provided with a feature chart containing features characterising only consonants.

4.4.2. Results

Predicted probabilities assigned to word-final singletons and geminates in the corpus were compared with native speaker' (dis)preferences of singletons over geminates. Predicted (model-assigned) probabilities of singletons and geminates do not match native speaker judgments: there is a somewhat stronger, but still weak positive correlation (r=0.31) between expected and observed probabilities. It is plotted as Figure 4.

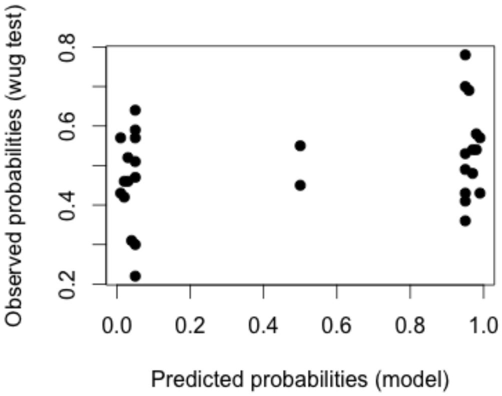

Figure 4: Simulation 3

Even though the correlation is somewhat stronger than in Simulation 2, the model is not more successful than the previous one. As is clearly illustrated by Figure 4, it is merely a pathological correlation, mostly involving points at both endpoints of the line, but not much along the entire *x* dimension.

4.5. A note on failed simulations

In the previous section, we tried different types of learning processes. After Simulation 1 - the first unsuccessful simulation - we attempted to restrict the learning process and draw the learner's attention to more and more specific and relevant information. Matching up native speakers' judgements with corpus frequencies did not prove to be successful in any of the simulations.

The lack of success in matching up the two sets of data can be attributed to various factors. One such factor is the fairly small number of monosyllables in the native Hungarian lexicon and the resulting (accidental) gaps. Because of that, people are not able to replicate fine-grained details (e.g. different vowel + consonant combinations, differences in place of articulation etc.) in their judgements on nonce words. However, they are able make broader generalisations, that is, they have intuitions about which consonant classes are more common as geminates.

4.6. Simulation 4

Based on the above discussion, a new, more restricted learning model is needed to account for native speakers' broad generalisations about geminate markedness.

4.6.1. Settings

The learner was tested on final consonants represented as consonant classes. Other parts of words were excluded and word-final consonants (both singletons and geminates) were organised into a list. Individual consonants were collapsed across consonant classes. Voiceless stops were transcribed as T or T:, voiced stops as D or D:, voiceless fricatives as S or S:, voiced fricatives as Z or Z:, voiceless affricates as TS or TS:, nasals as N or N:, and liquids as L or L:.[5]

4.6.2. Results

Predictions by the learner and native speaker judgements on (dis)preferences for singletons over geminates by each consonant class were compared. Figure 5 shows that native speakers' preferences

[5] Data containing voiced affricates and glides were not included in the training and testing data, for the following reasons: voiced affricates are always pronounced long in the relevant contexts, and not many recent loanwords contain glides.

match corpus frequencies fairly well: there is a strong positive correlation between wug test results and type frequencies (r=0.85). What is illustrated by Figure 5 is qualitatively similar to what we saw in Figure 1.

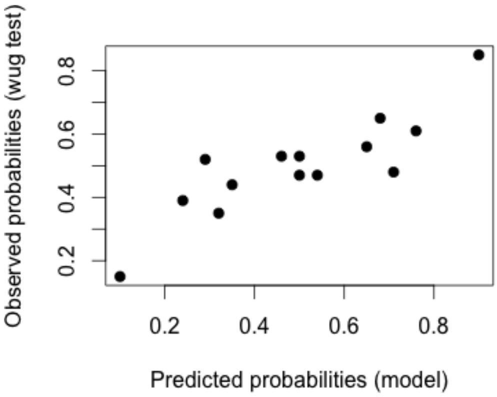

Figure 5: Simulation 4

Thus, we can see that the most successful model of humans' judgments is a learning model that is constrained to focus on the relative probabilities of singletons vs. geminates for broad natural classes of segments, rather than for specific consonants or vowel + consonant combinations.

5. Conclusions

The goal of this paper was to test the following three hypotheses: (1) Even native speakers of Hungarian (a language which allows all kinds of geminates) have some awareness of cross-linguistic geminate markedness; (2) These markedness asymmetries are also reflected in the native lexicon: the frequency distribution of geminates in the native phonology reflects patterns of universal markedness; (3) These patterns can be learnt from the native lexicon based on phonotactic generalisations.

Results of a nonce word well-formedness task confirmed Hypothesis 1: native speakers do have awareness of patterns of cross-linguistic geminate markedness, as they generally rate cross-linguistically less marked geminates higher than more marked ones. Their judgements also show similar preferences to what we can observe in Hungarian loan gemination processes.

Whether Hypothesis 2 has been confirmed by the type frequency distribution of geminates and singletons in the native Hungarian phonology is a question less straightforward to answer. Cross-linguistically less marked geminates are more commonly found in the corpus than more marked ones. However, it is not clear whether native speakers' preferences seen in nonce word judgements and loanword adaptation processes is a reflection of universal markedness or it directly comes from the native phonology. Also, the fact that the frequency distribution of geminates and singletons by consonant class does line up with cross-linguistic markedness patterns may indicate that gemination markedness is grounded in phonetics: native speakers have a knowledge of this, but this knowledge is less explicit in the case of native speakers of Hungarian (and native speakers of any language which allows all consonants to be geminated).

The confirmation of Hypothesis 3 is also complicated. Although it is possible to construct a learning model which mirrors native speakers' broad generalisations of gemination by consonant class, such a model has to be very much restricted. It is possible that native speakers can learn cross-linguistic geminate asymmetries, but only if their attention is drawn to details specific to geminate markedness. In particular, in the course of running simulations, the results only improved when geminate markedness was `dissociated' from the surrounding vowels, that is, when the model was trained in a way that VC co-occurrences do not matter.

The question arises whether VC co-occurrence is a more salient feature of the language than it appears to be at first sight and it does matter, but the learning process stops before it can learn all constraints relevant to VC co-occurrence restrictions. However, its is very unlikely that this is the case: the learner was run several times with and without restricting the number of constraints, gram sizes and O/E values, but the results turned out to be similar in each case.

As mentioned earlier, it is possible that, because of the small number of monosyllables, more fine grained details are ignored in the learning process and native speakers are forced to generalise in a coarser way. This is the reason why we developed a simplified model. However, it raises the question whether such simplifications are indeed necessary, or if we look at polysyllables as well, we will see a different pattern: that is, native speakers generalise over both monosyllables and polysyllables, and since they have a greater exposure to different VC sequences, they are able to make more fine-grained distinctions as well. Apart from polysyllables, token frequency distributions might also be worth considering, even though it is typically type frequency which plays a role in such processes and influence native speakers' judgements.

References

Berger, Adam L., Stephen A. Della Pietra & Vincent J. Della Pietra (1996). A Maximum Entropy approach to natural language processing. *Computational Linguistics 22:39-71.*

Della Pietra, Stephen A., Vincent J. Della Pietra & John D. Lafferty (1997). Including features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19:380-393.*

Eisner, Jason (2000). Review of Kager: "Optimality Theory". *Computational Linguistics, 26(2):286-290.*

Eisner, Jason (2001). Expectational semirings: Flexible EM for finite-state transducers. *In: Gertjan van Noord (ed.), Proceedings of the ESSLLI Workshop on Finite-State Methods in NLP (FSMNLP).*

Goldwater, Sharon & Mark Johnson (2003). Learning OT Constraint Rankings Using a Maximum Entropy Grammar. *In: Jennifer Spenader, Anders Eriksson and Östen Dahl (eds.), Proceedings of the Stockholm Workshop on Variation within Optimality Theory, 111-120. Stockholm: Stockholm University, Department of Linguistics.*

Halácsy, Péter, András Kornai, László Németh, András Rung, István Szakadát & Viktor Trón (2004). Creating open language resources for Hungarians. *Proceedings of the 4th International Conference on Language Resources and evaluation.*

Hayes, Bruce & Colin Wilson (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry.*

Jäger, Gerhard C. (2004). Maximum entropy models and stochastic Optimality Theory. *Rutgers Optimality Archive ROA-625.*

Jelinek, Frederick (1999). Statistical methods for speech recognition. *Cambridge, MA: MIT Press.*

Karvonen, Daniel (2009). The Emergence of the Unmarked in Finnish loanword phonology. *Paper presented at the 17th Manchester Phonology Meeting.*

Kawahara, Shigeto (2007). Sonorancy and geminacy. *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality III.*

Kertész, Zsuzsa (2006). Approaches to the phonological analysis of loanword adaptation. *The Even Yearbook 7, Department of English Linguistics, Eötvös Loránd University, Budapest.*

Klein, Dan & Chris Manning (2003). Maxent models, conditional estimation, and optimization, without the magic. *Tutorial presented at NAACL-03 and ACL-03.*

Kraehenmann, Astrid (2011). Initial geminates. *The Blackwell Companion to Phonology, Chapter 47, Wiley.*

Krishnamurti, Bhadriraju & John Peter Lucius Gwynn (1985). A Grammar of Modern Telugu. *Oxford University Press.*

Kubozono, Haruo, Junko Ito & Armin Mester (2008). Consonant gemination in Japanese loanword phonology: A phonological account. *Proceedings of the 18th International Congress of Linguists.*

Manning, Chris & Hinrich Schütze (1999). Foundations of statistical natural language processing. *Cambridge, MA: MIT Press.*

Morén, Bruce (1999). Distinctiveness, Coercion and Sonority: A Unified Theory of Weight. *PhD Dissertation, University of Maryland.*

Nádasdy, Ádám (1989). Consonant length in recent borrowings into Hungarian. *Acta Linguistica Hungarica* 39.

Passino, Diana (2004). Adaptation of loanwords and licensing strategies in Italian. *Paper presented at the 12th Manchester Phonology Meeting.*

Podesva, Robert (2002). Segmental constraints on geminates and their implications for typology. *The 76th Annual Meeting of the Linguistics Society of America.*

Rosenfeld, Ronald (1996). A maximum entropy approach to adaptive statistical language modeling. *Computer Speech and Language 10:187-228*.

Sridhar, Shikaripur N. (1990). Kannada. *New York: Routledge*.

Steriade, Donca (2004). Sources of markedness and why they matter. *GLOW, Markedness Workshop*.

Thurgood, Graham (1993). Geminates: A cross-linguistic examination. *In: Joel Ashmore Nevis, Gerald McMenamin & Graham Thurgood (eds.), Papers in Honor of Frederick H. Brengelman on the Occasion of the Twenty-Fifth Anniversary of the Department of Linguistics, California State University, Fresno*.

Proceedings of the 33rd West Coast Conference on Formal Linguistics

edited by
Kyeong-min Kim, Pocholo Umbal,
Trevor Block, Queenie Chan,
Tanie Cheng, Kelli Finney, Mara Katz,
Sophie Nickel-Thompson, and Lisa Shorten

Cascadilla Proceedings Project   Somerville, MA   2016

Copyright information

ISBN 978-1-57473-469-0 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Magyar, Lilla. 2016. Are Universal Markedness Hierarchies Learnable from the Lexicon? The Case of Gemination in Hungarian. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, ed. Kyeong-min Kim et al., 287-298. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3248.