Contextual Constraints on Geminates: The Case of Polish*

BOŻENA PAJĄK
University of California, San Diego

0. Introduction
In this paper I argue from typological and perceptual evidence that the constraint against geminates (*Gem) should be split into constraints that incorporate contextual information (word position & adjacent segments). I show that splitting *Gem accurately captures the distribution and conspiratorial behavior of geminates in Polish.

1. Background
Geminates can be described as long consonant. Cross-linguistically, they are on average between one-and-a-half and three times as long as singletons (Ladefoged and Maddieson 1996). Geminates are often used contrastively in languages, as illustrated by the examples in (1).

(1) Italian: *bello – belo (‘beautiful’ – ‘I bleat’)
    Finnish: *takka – taka- (‘fireplace’ – ‘back’)

While geminates can vary greatly in the way they are represented phonologically, the discussion in this paper includes all geminates regardless of their exact structural representation (e.g., consonants with two timing slots, a single mora projection, two adjacent identical segments, etc.; see e.g. Hume, Muller, and Engelenhoven 1997, Davis 1999, Topintzi 2008).

In Optimality Theory (OT), the commonly used constraint against geminates is *Gem (Rose 2000). There have been proposals to split *Gem into a family of constraints targeting particular segmental types of geminates, as shown in (2) and (3), based on both typological and perceptual evidence. The main idea was to account for differences between, for example, geminate obstruents and geminate

* I would like to thank Eric Baković, Lucien Carroll, Noah Girgis, Alex del Giudice, Cynthia Kilpatrick, Sharon Rose, and the audience of BLS 35 for valuable comments and suggestions.
1 The examples are from on-line dictionaries: http://www.wordreference.com/iten (for Italian) and http://www._ncd.com/ (for Finnish).
sonorants. The latter are typologically less common and perceptually less salient (at least in the intervocalic environment), and thus the constraints against geminate sonorants are assumed to outrank the constraints against geminate obstruents (see e.g. Kawahara 2007).

\[(2) \quad \text{*GemGutt} \gg \text{*GemSon, *GemFric, *GemVoicedObs} \]
\[\text{*GemGlide} \gg \text{*GemLiquid} \gg \text{*GemNasal} \]
(Podesva 2002)

\[(3) \quad \text{*GemGlide} \gg \text{*GemLiquid} \gg \text{*GemNasal} \gg \text{*GemObs} \]
(Kawahara 2007)

In this paper I am concerned not with the segmental composition of geminates, but with the context in which they appear, where by context I mean their word position and adjacent segments. Previous work has shown that segmental context is an important property that often needs to be taken into account in the phonological analyses of geminates (see e.g. Muller 1999, McCrory 2004). Adjacency to vowels appears to be especially significant: typological evidence shows that geminates are most common intervocally, and most rare when not adjacent to any vowel (Thurgood 1993, Muller 2001; plus an informal survey of 40 languages with geminates). This typological fact correlates with perceptual evidence (at least as tested for obstruents): intervocalic singleton-geminate contrasts are the most perceptible, and non-vowel-adjacent singleton-geminate contrasts are the least perceptible (Pajak 2009; see also McCrory 2004, Dmitrieva 2009).

2. Proposal
Vowel adjacency thus constitutes an important property that helps define common and uncommon geminate contexts. This property can be incorporated into phonological theory by re-defining *Gem as a family of constraints that target geminates in different contexts. This is analogous to the proposal of splitting *Gem into a family of segmental constraints, as discussed in §1. Informal definitions of the proposed contextual constraints on geminates are shown in (4).

\[(4) \quad \text{Informal definitions of contextual constraints on geminates} \]
\[\text{*Gem/V_V} \quad \text{Geminates flanked by vowels are not allowed (‘no intervocalic geminates’).} \]
\[\text{*Gem/1VA} \quad \text{Geminates adjacent to exactly one vowel are not allowed (‘no single vowel-adjacent (1VA) geminates’).} \]
\[\text{*Gem/NVA} \quad \text{Geminates not adjacent to any vowel are not allowed (‘no non-vowel-adjacent (NVA) geminates’).} \]
These contextual constraints may need to be more specific than defined here, incorporating information about word position (e.g., *GEM/#GGV, *GEM/VGG#, etc.) or combining with segmental-type constraints (e.g., *GEMObs/NVA). For present purposes, however, such considerations are left as open questions depending on further evidence.

A universal ranking of these contextual constraints can be established based on the typological and perceptual facts noted earlier, as shown in (5). The constraint against non-vowel-adjacent geminates is ranked the highest, while the constraint against intervocalic geminates is ranked the lowest. This hierarchy of constraints predicts certain implicational universals. Namely, the presence of non-vowel-adjacent geminates in a language implies the presence of one-sided vowel-adjacent geminates, which in turn implies the presence of intervocalic geminates. This is consistent with Thurgood’s (1993) conclusion that if a given language allows geminates in any other environment than flanked by vowels, it also necessarily allows them intervocally.

(5) Universal ranking of contextual constraints on geminates

\[
\begin{align*}
\text{*GEM/NVA} & >> \text{*GEM/IVA} & >> & \text{*GEM/V_V} \\
#GGC, CGG#, CGGC & #GGV, VGG#, CGGV, VGGC & VGGV
\end{align*}
\]

The only potential counterexamples to this universal ranking are languages which seem to allow word-initial geminates but not medial intervocalic ones, such as Pattani Malay, Iban, Sa’ban (Austronesian), or Nhaheun (Austro-Asiatic) (Blust 1995, 2007, Muller 2001). However, there are independent diachronic factors responsible for the apparent exceptionality of these cases. Initial geminates (or geminates in general) in many Austronesian languages (such as Pattani Malay or Iban) were created by a widespread diachronic process of vowel syncope between two identical consonants, which was motivated by a preference for disyllabic canonical shape (Blust 2007). In Sa’ban, initial geminates arose through a general process of unstressed vowel deletion in penultimate syllables (Blust 2001, 2007). In Nhaheun, on the other hand, most words are monosyllabic, which precludes any generalization concerning possible medial geminates (Muller 2001).

3. The Case of Polish

The proposed contextual constraints in (4) and their ranking in (5) are central to the account of the overall distribution of geminates in Polish, which is shown in the analysis developed in this section.

3.1. Geminates in Polish

Polish has a phonemic distinction between singleton and geminate consonants: e.g., [buda] ‘kennel’ and [budda] ‘Buddha’. There are examples of both ‘true’ geminates, which are underlyingly long (mostly borrowings from other languages), and of ‘fake’ geminates, which are derived through certain morphologi-

Geminates in Polish behave fairly typically when compared to other languages in that they are mainly found intervocally, as shown in (6).

(6) **Intervocalic geminate consonants**

a. **Sonorants**

- fonantanna ‘fountain’
- gamma ‘gamma’
- dzeñnik ‘gazette’

b. **Obstruents**

- getto ‘ghetto’
- lekkɔ ‘lightly’
- óddać ‘to give back’

When there is the potential to create a non-intervocalic geminate in Polish (e.g., via affixation), one of the consonants of the would-be geminate is deleted (Rubach and Booij 1990), as shown in (7)-(10). I refer to this deletion process here as *degemination*.

In (7a), single vowel-adjacent geminates could be created by adding the suffix -ni to stems ending with Cn. However, degemination applies instead. The comparison examples in (7b) show that deletion does not occur when the stem ends with different consonants. Furthermore, it is even possible to create a geminate, as the example of ‘sleep’/‘sleepy’ illustrates, as long as it is intervocalic.

(7) a. **Degemination postconsonantally**

- pčkn-o ‘beauty’
- kupn-o ‘purchase’

b. **No deletion**

- vôd-a ‘water’
- včtš-e ‘wind’ (Loc.)
- sen ‘sleep’

The same process can be observed in (8a), where preconsonantal geminates are avoided. Note that degemination applies equally to a monomorphemic stem-final geminate [I] and to a potential ‘fake’ geminate [s] that would be created across an affix boundary. What these two cases have in common is the fact that a geminate is banned due to the presence of an adjacent following consonant. Again, there is no deletion in any other cases, as shown in the comparison examples in (8b).

(8) a. **Degemination preconsonantally**

- sevill-a ‘Seville’
- frantšus ‘Frenchman’

272
b. No deletion

ekfador  ‘Ecuador’  ekfador-ski  ‘Ecuadorian’
serp  ‘Serb’  serp-ski  ‘Serbian’

Degemination also occurs word-finally, as shown in (9a). Stem-final geminates surface when followed by a vowel suffix, but degeminate when no (or zero) suffix is present on the stem. The comparison examples in (9b) show that the deletion is not enforced by a ban on word-final coda clusters.

(9)  

a. Degemination word-finally

fɔntann-i  ‘fountains’ (Nom.)  fɔntan (Gen.)  *fɔntann
flɔtill-ɛ  ‘fleets’ (Nom.)  flɔtıl (Gen.)  *flɔtill
lass-a  ‘lassoes’ (Nom.)  las (Gen.)  *lass

b. No deletion of final cluster

palm-i  ‘palms’ (Nom.)  palm (Gen.)
r uzg-i  ‘rods’ (Nom.)  rusk (Gen.)
vaxt-i  ‘duty watches’ (Nom.)  v axt (Gen.)

Degemination also optionally applies in the same segmental contexts at clitic and word boundaries (Sawicka 1995:153), as shown in (10a). Although consonant-adjacent geminates are tolerated in these cases, the optional repair available in this context (i.e., degemination) is the same as in all other potential single vowel-adjacent geminates.

(10)  

a. Optional degemination

bes+strɔnni  ~  be+strɔnni  ‘impartial’
raż+zwatʃɛtʃɛ  ~  r aż+zwatʃɛtʃɛ  ‘to enrage’
k ašk##kazdɪ  ~  kas##kazdɪ  ‘every helmet’

b. No deletion

bes+pwɪʃɔvi  *be+pwɪʃɔvi  ‘sexless’
raż+gaʃɛ  *raż+gaʃɛ  ‘to enrage’
raż+sadziʃɛ  *r aż+sa dziʃɛ  ‘to blow up’

There is, however, one case in which degemination is blocked: word-initial geminates can be formed with monoconsonantal proclitics /v/ and /z/, as illustrated in (11) (voicing assimilation in obstruent clusters is obligatory in Polish; e.g. Bethin 1992). Note that monoconsonantal proclitics differ from other clitics in that they cannot be syllabified separately from its host (Rubach and Booij 1990, Sawicka 1995, Rochoń 2000). This is in contrast to longer proclitics, as in (10), in which the final consonant is never resyllabified to form part of an onset but always remains in coda position.
Bożena Pająk

(11) **Vowel-adjacent initial geminates: no degemination**

\[
\begin{align*}
/v+/vɔzίtɛ & \rightarrow v+vɔzίtɛ \quad \text{‘to carry in’} \quad *O+vɔzίtɛ \\
/v+/fɔtɛlʊ & \rightarrow f+fɔtɛlʊ \quad \text{‘in an armchair’} \quad *O+fɔtɛlʊ \\
/z+/zɛbɛm & \rightarrow z+zɛbɛm \quad \text{‘with a tooth’} \quad *O+zɛbɛm \\
/z+/sunʒɛ & \rightarrow s+sunʒɛ \quad \text{‘to slip down’} \quad *O+sunʒɛ
\end{align*}
\]

Polish also has four monomorphemic words with initial geminates – three of them affricates – plus a few more forms derived from these: [sɔɑtɛ] ‘to suck’, [tʃtʃi] ‘empty’, [dʒtʃɔʃa] ‘earthworm’, and [dʒtʃiʃtʃi] ‘rainy’. Due to this limited number of examples, I conclude that they are simply exceptions to a ban on monomorphemic word-initial geminates in Polish. This conclusion receives some support from the fact that – in contrast to medial affricate geminates – initial affricate geminates are always pronounced as two separate consonants (Dunaj 1985), which casts some doubt on whether they are in fact geminates.

The final piece of data concerns the fact that word-initial geminates are only tolerated when adjacent to a vowel. Potential preconsonantal initial geminates formed with monoconsonantal proclitics are instead repaired by vowel epenthesis, as shown in (12a). The comparison examples in (12b) show that epenthesis does not apply to simply break a cluster because Polish allows very complex onset clusters.

(12) a. **Potential consonant-adjacent initial geminates: vowel epenthesis**

\[
\begin{align*}
/v+/vzɛcɲu & \rightarrow vz+vzɛcɲu \quad *v+vzɛcɲu, *O+vzɛcɲu \\
/v+/frunʒɛ & \rightarrow vz+frunʒɛ \quad *frunʒɛ, *O+frunʒɛ \\
/z+/znakəm & \rightarrow z+znaκəm \quad *z+znaκəm, *O+znaκəm \\
/z+/stazɛtɛ+cɛ & \rightarrow z+zstazɛtɛ+cɛ \quad *z+zstazɛtɛ+cɛ, *O+zstazɛtɛ+cɛ
\end{align*}
\]

b. **Potential consonant-adjacent non-geminate cluster: no epenthesis**

\[
\begin{align*}
z+bz̥ɛk′em & \quad \text{‘with a plunk’} \quad *z+ɛbz̥ɛk′em \\
s+pʃʃɔwɔ & \quad \text{‘with a bee’} \quad *z+ɛpʃʃɔwɔ
\end{align*}
\]

There is only one exception to this generalization: epenthesis does not apply to the word [s+stʃpɪtɛ] ‘to descend’ (plus other paradigmatic variants), a word that has fallen out of use and is used almost exclusively in rote religious contexts (as in ‘God descended on earth’).

Finally, the reader might be familiar with the fact that Polish proclitics are often assumed to end in an underlying abstract vowel called a ‘yer’, which is vocalized when followed by an unvocalized underlying yer in the following syllable (e.g., Szpyra 1992). On the surface, yer vocalization appears identical to the process of vowel epenthesis described here. However, the discussed vowel epenthesis is completely independent from the process of yer vocalization since
there are no underlying yers in (the first syllable of) the stems like the ones shown in (12) (see e.g., Rubach 1977, 1985).

The table in (13) summarizes the distribution of geminates in Polish. Intervocalic geminates are allowed, and so are single vowel-adjacent initial geminates when formed with monoconsonantal proclitics. All other would-be single vowel-adjacent geminates undergo degemination, and non-vowel-adjacent geminates – in the one context where they could potentially be created – are repaired by vowel epenthesis.

(13) **Distribution of geminates in Polish**

<table>
<thead>
<tr>
<th></th>
<th>VGGV</th>
<th>allowed</th>
</tr>
</thead>
<tbody>
<tr>
<td>intervocalic geminates</td>
<td>#G+GV</td>
<td></td>
</tr>
<tr>
<td>single vowel-adjacent geminates</td>
<td>VGGC, CGGV, VGG#</td>
<td>degemination</td>
</tr>
<tr>
<td>non-vowel-adjacent geminates</td>
<td>#G+GC</td>
<td>epenthesis</td>
</tr>
</tbody>
</table>

3.2. **Analysis**

I argue that the behavior of geminates in Polish constitutes a classic case of a conspiracy (Kisseberth 1970, Pater 1999). Two processes – deletion and epenthesis – conspire to avoid non-intervocalic geminates. Only intervocalic geminates seem to be freely allowed in the language. Whenever a geminate would be expected to surface in a non-intervocalic context due to morphological concatenation, degemination takes place instead. However, degemination is blocked whenever – as I assume – it would lead to the loss of the entire proclitic. In these cases word-initial geminates are either tolerated (when prevocalic, or single vowel-adjacent) or repaired by vowel epenthesis (when preconsonantal, or non-vowel adjacent). (Casali (1997:506ff) discusses similar cases in which the result of an otherwise expected vowel deletion process is blocked just in case an entire morpheme would be sacrificed.)

This pattern can be straightforwardly accounted for with the proposed contextual constraints on geminates, defined in (4). Additional constraints necessary for the analysis are shown in (14).

(14) **Informal definitions of additional constrains**

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEP(V)</td>
<td>No vowel epenthesis.</td>
</tr>
<tr>
<td>MAX(C)</td>
<td>No consonant deletion (no degemination).</td>
</tr>
<tr>
<td>REALIZE MOR(PHEME)</td>
<td>An input morpheme must have some phonological exponent in the output (e.g., Kurisu 2001).</td>
</tr>
</tbody>
</table>

The full OT analysis is provided below. In the tableau in (15), the candidate with an intervocalic geminate (a) surfaces as optimal because other candidates are eliminated by higher-ranked constraints. The degeminated candidate (b) violates
Bożena Pająk

MAX(C), and the candidate with epenthesis (c) violates Dep(V). Therefore, the constraints Dep(V) and MAX(C) must dominate NoGem/V\_V.

(15) **Intervocalic geminates**

<table>
<thead>
<tr>
<th>/lassɔ/</th>
<th>Dep(V)</th>
<th>MAX(C)</th>
<th>*Gem/V_V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → lassɔ</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. lassɔ</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. lasesɔ</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The tableau in (16) shows how degemination is enforced in order to avoid a single vowel-adjacent geminate. The degeminated candidate (b) wins because the faithful candidate (a) is eliminated by the higher-ranked constraint NoGem/1VA. Note that in this case the candidates with epenthesis (in any position) (c-d) are also not possible. This justifies ranking both NoGem/1VA and Dep(V) above MAX(C).

(16) **Degemination**

<table>
<thead>
<tr>
<th>/sevill-ski/</th>
<th>Dep(V)</th>
<th>*Gem/1VA</th>
<th>MAX(C)</th>
<th>*Gem/V_V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. sevillski</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. → sevilski</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. sevilelski</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. sevilleski</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

The tableau in (17) illustrates how degemination is blocked just in case it would lead to the complete loss of a proclitic. The candidate with an initial geminate (a) surfaces as optimal despite violating the constraint NoGem/1VA because the degeminated candidate (b) is eliminated by REALMOR, while the candidates with epenthesis (c-d) are again eliminated by Dep(V). Thus, the correct result is obtained when both REALMOR and Dep(V) outrank NoGem/1VA.

(17) **Initial geminates**

<table>
<thead>
<tr>
<th>/v+ɔziʨ/</th>
<th>REALMOR</th>
<th>Dep(V)</th>
<th>*Gem/1VA</th>
<th>MAX(C)</th>
<th>*Gem/V_V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → ʋvɔziʨ</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ʋɔziʨ</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. ʋɛvɔziʨ</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ɛvʋɔziʨ</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Contextual Constraints on Geminates

The last case of interest concerns the situation in which a non-vowel-adjacent geminate is avoided through vowel epenthesis. This is shown in the tableau in (18). The candidate with epenthesis (c) is optimal because the faithful candidate with a non-vowel-adjacent geminate (a) is eliminated by NoGem/NVA, and the degeminated candidate (b) is eliminated by RealMOR. The second-best repair in this case is epenthesis. In order to obtain this result, both NoGem/NVA and RealMOR have to be ranked above Dep(V). Note also that epenthesizing a vowel immediately after the clitic is optimal because any other epenthesis location (as in (d) and (e)) incurs a violation of NoGem/1VA in addition to violating Dep(V).

(18) Epenthesis

<table>
<thead>
<tr>
<th>/v+vzęgnu /</th>
<th>Gem/NVA</th>
<th>RealMOR</th>
<th>Dep(V)</th>
<th>Gem/1VA</th>
<th>Max(C)</th>
<th>Gem/V_V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vvvzęgnu</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. vzęgnu</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. → vεvzęgnu</td>
<td></td>
<td></td>
<td>!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. εvzęgnu</td>
<td></td>
<td></td>
<td>*</td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. vεvzęgnu</td>
<td></td>
<td></td>
<td>*</td>
<td>!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The summary of the constraint ranking that accounts for the distribution of geminates in Polish is provided in (19) (overleaf). The non-vowel-adjacent geminates are disallowed due to the high-ranked constraint *Gem/NVA. The repair of vowel epenthesis is enforced by RealMOR which crucially outranks Dep(V). The single-vowel-adjacent geminates undergo degemination, which is assured by ranking *Gem/1VA above Max(C). The tolerance for single-vowel-adjacent geminates (created with proclitics) in the word-initial position is again enforced by high-ranked RealMOR. Finally, intervocalic geminates are freely allowed due to the low ranking of *Gem/V_V.

4. Conclusion

I have shown that context (defined here as word position & adjacent segments) is an important characteristic of geminates. Based on typological and perceptual evidence, I argued that the constraint against geminates, *Gem, should be split into at least three general contextual constraints: *Gem/NVA >> *Gem/1VA >> *Gem/V_V. Finally, I showed how these constraints correctly account for the distribution of geminates in Polish.
Bożena Pająk

(19) Constraint ranking responsible for the distribution of geminates in Polish

<table>
<thead>
<tr>
<th>non-vowel-adjacent geminates</th>
<th>#G+GC</th>
<th>epenthes</th>
</tr>
</thead>
</table>
| single vowel-adjacent geminates | VGGC  
|                               | CGGV  
|                               | VGG#  
| allowed                       |       
| intervocalic geminates        | VGGV  

References


Muller, Jennifer. 2001. The phonology and phonetics of word-initial geminates. Ph.D. diss., Ohio State University.
Bożena Pająk


Bożena Pająk
University of California, San Diego
Department of Linguistics
9500 Gilman Drive #108
La Jolla, CA 92093-0108

bpajak@ling.ucsd.edu