The productivity of Dutch diminutives

Rosita van Tuijl & Peter Coopmans
Utrecht University

This study reports on an investigation of the productivity of the Dutch diminutive paradigm, which gives rise to five suffixal allomorphs, against the background of Yang's (2016) Tolerance Principle. It shows how, by studying the frequency of the allomorphs and the environments in which they occur, we can use the Tolerance Principle to determine if a productive rule system can be found for the Dutch diminutives. In doing so, we also describe how we collected the necessary data by setting up a corpus study on Dutch diminutives, including a specific one with data from child directed speech.

Keywords: Dutch diminutives, tolerance principle, computational linguistics, acquisition

1.   Introduction

In this paper we aim to test if Yang's (2016) Tolerance Principle (TP) can be used to predict a productive rule system for the various allomorphic realisations of the Dutch diminutive suffix. The TP will help us determine from a quantitative angle which of these suffixation rules could be productive, how many exceptions they allow, and how this contributes to a feasible learnability scenario for the Dutch language learner. We will test if the TP is a good predictor of productivity if we limit the data the child builds on to the features whose relevance has been convincingly shown in morpho-phonological studies of Dutch diminutives. In addition, we will use developmental research as another measure of productivity to guide our search.

In Section 2, we present a description of the Dutch diminutive paradigm, relying on relevant previous research on diminutive suffixation, highlighting some proposed theoretical generalisations and acquisition findings available. Section 3 starts out with a brief outline of the reasoning behind the TP and then describes the corpus studies performed to gather the data necessary to invoke the TP in determining the system's productivity. The analysis is given in Section 4, with a

discussion of its outcomes and further considerations. Section 5 contains some concluding remarks.

2.   Dutch diminutives

2.1   Empirical background

In Dutch, like in many other languages, diminutives are formed by attaching a suffix to the stem, the base form of a lexeme. We observe five allomorphic realisations of this diminutive suffix: *-je, -tje, -kje, -pje* and *-etje.* Which allomorph is used is determined by the phonological features of the stem. Previous research (Section 2.2) has carefully determined that the most important features in this process are: (i) the sonority of the stem's ending, (ii) the length of the preceding vowel and (iii) the stem's stress pattern. Here are some basic examples showing the effect of these features on the allomorphic realisations.

i.   The role of sonority is visible in examples like *boek* 'book', whose final consonant /k/ is an obstruent like [p, t, f, s]. Its diminutive form is *boek-je* 'book-DIM'. We similarly have *roos-je* 'rose-DIM' and *taart-je* 'cake-DIM'. Sonorant endings, provided by final vowels or consonants like [m, j, l, n, ŋ, r] give rise to diminutive forms with the suffixal forms *-tje, -pje* or *-kje*.
ii. The relevance of vowel length is visible in such cases as *zoon* 'son' vs *zon* 'sun', where the form with the long vowel is diminutivised as *zoon-tje* 'son-DIM', the latter as *zonn-etje* 'sun-DIM'.
iii. The role of stress in the correct choice of allomorph revolves around the question whether the root word has a stressed penultimate syllable or not. A word like *woning* 'house' has penultimate stress while any monosyllabic word or words like *wa-gon* 'wagon', which has final stress, do not. *Woning*'s diminutive form is *wonin-kje*, but *wagon* 'wagon-DIM' ends up as *wagonn-etje*.

We should also mention the occurrence of homorganic variance. The allomorphs *-kje* or *-pje* can be treated as homorganic variants of *-tje*, in accordance with place assimilation conditions (Kooij & van Oostendorp, 2003). When a noun ends in /m/, the *-pje* variant is selected, when it ends in /ŋ/, *-kje* is selected.

Selection of the diminutive allomorphs can be descriptively captured by the rules in (1–5) as shown by Linke (2020), with some illustrations.

(1) If stem has obstruent ending, select *-je.*
     roos 'rose'   *roosje*
     drop 'licorice' *dropje*

(2) If stem has sonorant ending preceded by short vowel and no penultimate
     stress, select *-etje.*
     zin   'sentence' *zinnetje*
     ballon 'balloon' *ballonnetje*

(3) If stem has stem final /m/, select *-pje.*
     bodem 'bottom' *bodempje*
     raam   'window' *raampje*

(4) If stem has stem final /ŋ/, select *-kje.*
     koning 'king' *koninkje*

(5) In all other cases, select *-tje.*
     dame 'lady' *dametje*
     zoen 'kiss' *zoentje*

2.2   Theoretical accounts

In some theoretical studies it has been argued that the Dutch diminutive has a
default suffix form from which the other allomorphs can be derived. The two most
popular candidates for this default suffix are *-tje* and *-je*, which is not entirely sur-
prising as these are the two forms that occur with the highest frequency.

  Kooij & van Oostendorp (2003) argued for *-tje* as the default form. The reason
for doing so is that *-tje* occurs after stems that end in a glide or in a vowel. If the
suffix follows an obstruent, /t/ will be omitted, giving *-je* as the suffixal form. The
other allomorphs will then be construed through place assimilation (*-kje, -pje*)
and schwa-insertion (*-etje*).

  Booij (1999) considered *-tje* the default form as well. His theory was similar
to Kooij & van Oostendorp's, but instead of place assimilation for the *-pje* and
-kje allomorphs Booij argued that in these cases /t/ is deleted and /k/ and /p/ are
inserted as homorganic stops.

  Van de Weijer (2002) also proposed deriving the other allomorphs from the
-tje suffix and he did so in an Optimality Theoretical (OT) framework by such
universal constraints as Place Assimilation and Faithfulness to root over affix.

  Trommelen's (1984) original study also argued for *-tje* as the default. One of
her central claims was that the selection of the allomorph can be entirely deter-
mined by properties of the final syllable. This stands in contrast with the descrip-

tions given by most other researchers, because it excludes from the condition any reference to [+/− stress] nature of the penultimate syllable (see (2)).

In contrast to the studies above, Huber (2005) advocated taking *-je* as the default form, with the claim that this underlying form is turned into 'something palatal' whenever it is intervocalic (*zee-t-je* 'sea-DIM') or when it is preceded by a stem with a sonorant ending *(zoen-t-je* 'kiss-DIM')*. The *-pje* and *-kje* allomorphs arise as instances where /p/ and /k/ are ephenthetic stops.

An alternative account for selection of the proper allomorph was presented by Van der Hulst (2008), for whom the crucial factor is not the segmental properties of the stem but rather its prosodic structural organisation – specifically with sensitivity to the branching structure at the foot level -, which determines choice of either the allomorphic variant *-etje* or *-tje* (the latter further showing up as *-kje*, *-pje* due to assimilation).

2.3   Acquisition findings

On top of these theoretical proposals as to which suffix is the best candidate for default status and, in fact, whether there should be one to begin with, we can also turn to experimental studies to find out whether there is a base form with the others as derived allomorphic realisations.

Den Os & Harder (1987) administered a wug-test (Berko, 1958) to 4–12 year old Dutch children, and to an adult control group. They found a clear order of acquisition: *-je* is acquired first, closely followed by *-tje* and *-pje*. The children didn't fully acquire *-kje* until age 12, and within the experiment, they never fully mastered *-etje*. Interestingly and very relevant to our investigation, as will be discussed in the analysis, even the adult control group made mistakes with nonce words that should get the *-etje* allomorph.

Gillis (1997) studied diminutive acquisiton in a single child between 1;5–2;5 years. The child started producing diminutive forms at the age of 1;7. Gillis also determined to what extent the child had acquired the semantic relation between the diminutive suffix and the smallness of the objects referred to and noted that the child did not yet seem to understand the pragmatic use of diminutives at the moment of the final tests.

Boersma (2018) conducted both a judgement and a production test with Dutch adults and 5-to-10 year olds. Like Den Os & Harder (1987), Boersma found that the adults performed in conformity with the descriptive rules in (1–5) above, except for *-etje*. When diminutivising nonce nouns that should select this allomorph, the adults only gave target answers 43% of the time. Instead of *-etje,* they chose *-tje* (or *-kje/-pje*, in accordance with place assimilation). So, with

nonce words like *gol* and *vom*, whose expected diminutives would be *goll-etje* and *vomm-etje* (by rule (2)), many participants responded with *goltje* and *vompje*.

Boersma also found that not all her adult participants had a fixed strategy when it comes to novel nouns that should get *-etje*. Only 5% of them actually consistently chose *-etje*, 23% consistently chose either *-tje, -kje* or *-pje,* and a majority of 72% used both of these approaches interchangeably. For Boersma these results suggest that the nouns diminutivised with *-etje* are in fact lexicalised forms, because the adults do not seem to have a set strategy for diminutivising this particular set of words.

The results of the production and judgement tasks administered to the children show improvement by age in their application of the required diminutive suffixation rules, except for *-etje* and *-kje*. When these two allomorphs are involved, the children's accuracy on real nouns stays about the same in the different age groups. With nonsense nouns their accuracy on the *-etje* allomorph actually becomes worse as the children get older. More generally, it can be concluded that the older the children, the more adultlike they become in their use of *-etje*. This result further suggests that the nouns diminutivised by *-etje* are lexicalised and not derived by a productive rule.

We end our overview by concluding that the issue of whether there is a default form of the Dutch diminutive is far from settled. Although most researchers have either argued for *-je* or *-tje* as the default, there is no clear consensus on the nature of the underlying system. We do observe an agreed upon order of acquisition of the five allomorphs: *-je < -tje < -pje < -kje < -etje*. The allomorph with which Dutch-learning children struggle most is *-etje,* which even turns out to be problematic for adults when confronted with non-existing words.

We will next turn to the Tolerance Principle and show how Dutch diminutives provide a fertile testing ground in our search for the productivity of the underlying rules.

3.   Quantifying the relevant diminutive data distribution

3.1   The tolerance principle

Yang (2016) proposed the TP as a model for the way we distinguish the core of paradigms from the periphery. It helps language learners, and linguists as well, in finding the rules that define a paradigm and the exceptions to these rules. When acquiring language, we optimise our model of a paradigm by looking for patterns or rules, to make generalisations. The TP is a way to determine if any of these rules can be accepted as productive, or not.

The TP states that a productive rule can only have a limited number of exceptions. This number has to stay below a threshold, which itself is determined by the overall number of lexical items in the paradigm. If the number of exceptions exceeds the threshold, the learner has to revise her rule. The mathematical equation determining this threshold is given in (6).

(6)   Tolerance Principle
      Let *R* be a rule applicable to *N* items, of which *e* are exceptions. *R* is productive
      if and only iff
      $e \leq \theta_N$ *where* $\theta_N := N / ln\ N$ (Yang, 2016, p. 64)

The principle comes with a step-by-step routine for its application. First a rule *R* is obtained. The search for R itself is an instance of inductive learning that must proceed conservatively, which Yang (p. 42) illustrates with reference to the Yip-Sussman (1997) model. Rule *R* has a certain structural description and a specific set of lexical items it applies to. The next steps involve counting the number of lexical items that fit this structural description (*N*) and counting the items within this set that are exceptions to the rule (*e*). With this data the threshold can be calculated and the child learner can decide if the rule is productive or not. Crucially, the data with which such a calculation is performed should preferably come from child-directed speech (CDS), as this is the actual input a child relies on when acquiring a language.

For every set of items the TP can be used to determine how many exceptions a rule over this set can tolerate. For example, a set of 9 items can tolerate $\theta_9 = 4$ (rounded down) exceptions, a set of 24 items can handle $\theta_{24} = 7$ and a set of 300 can handle $\theta_{300} = 52$ exceptions.

The model example Yang uses to illustrate the TP is English past tense formation. Here there is one overwhelming rule: 'if past tense, add *-d*', which can be calculated as productive over the whole set of English verbs. If we look at this entire set and apply the TP, the number of irregular verbs stays below the calculated threshold.

Not all paradigms are like this, i.e., in having one overwhelming rule that covers the whole system. German plural formation is based on a complicated system in which none of the five possible plural markers has a frequency high enough to produce a rule that tolerates all the other markers as exceptions. So, with German pluralisation a system of multiple rules is necessary instead (see Yang, 2016, p. 122–136). In order to achieve such a rule system, he proposes the Maximize Productivity Principle (MPP), which encourages the learner to look for different rules if a hypothesis over the whole set *N* does not work out. This holds particularly for rules that divide *N* into subsets.

(7)   Pursue rules that maximize productivity.

If no rule can be found that is productive over the whole set of German nouns, the learner will continue on a search for subsets that might have a productive rule. Yang (2016) shows how this can be done successfully for German plurals, providing a fruitful model for our search into the appropriate rule system for Dutch diminutives. However, before we can attempt such an analysis to our case study, we need the proper data to extract the frequencies of the five allomorphs.

3.2   Corpus research: Determining allomorphs

We will describe here how we made a collection of Dutch diminutives from a large online corpus. For this goal we used two online platforms: GrETEL (Augustinus et al., 2014) and PaQu (Odijk, 2015). Both platforms can be used to search online Dutch annotated corpora and treebanks. Here we used GrETEL to make a detailed XPath query for diminutives that was then used in PaQu to select all the diminutives from all available Dutch corpora.

After collecting the diminutive forms, the next step was to determine which allomorph each word receives. The simplest approach to this task would be to compare the end of the diminutive with the respective suffix and check if they match (e.g., the end of *boekje* matches with *-je*, so it receives the *-je* allomorph). However, due to the similarity of the different suffixal forms, this approach would often result in false positives.

For example, if we built an algorithm that started by comparing all diminutive endings with *-je*, all would return a match because the other four allomorphs also contain *-je*. A solution could be to start out with the longer allomorphs, so none of those would be miscategorised as having *-je*, and then keep *-je* as the default for the remaining set. However, as is shown in (8), this would also result in incorrect categorisations.

(8)   Problematic categorisations

Diminutive	Stem	Selected	Correct
a. *boekje* ('book-DIM')	*boek*	*-kje*	*-je*
b. *hertje* ('deer-DIM')	*hert*	*-tje*	*-je*
c. *dametje* ('lady-DIM')	*dame*	*-etje*	*-tje*
d. *aapje* ('monkey-DIM')	*aap*	*-pje*	*-je*

These examples show that a simple search that matches the diminutive with a suffix is not feasible. In order to categorise the diminutives correctly, we specifically need the stem of each diminutive. Luckily, PaQu can be used to extract the stem of each noun as well. The proper division was made by comparing each diminutive

with its respective stem. For example, if the stem is *boek* ('book') and the diminutive is *boekje* ('book-DIM'), the suffix is found by searching *boekje* for *boek* and returning what is left over. This would be *je*, searching *boompje* for *boom* would return *pje* etc.

In this way almost all diminutives can be matched with a suffix and the frequencies of the allomorphs can be counted (Table 1).[1]

Table 1. Type, token frequencies of diminutive allomorphs from PaQu (excl. CHILDES corpus)

Allomorph	Type		Token
-je	1824	(56%)	511973
-tje	964	(30%)	155114
-kje	113	(3%)	625
-pje	82	(3%)	10436
-etje	234	(7%)	44362
Rest	28	(1%)	135
Total	**3245**	**(100%)**	**722645**

3.3   Corpus research: Feature selection

Now that we have been able to determine the actual corresponding suffixes for the diminutives, we can start dividing all the nouns into categories based on the phonological features in Section 2.1.: sonority, vowel length and the stress pattern.

By dividing the collected noun stems into subsets based on these features, we can observe the frequencies of the five allomorphs within these subsets. For example, according to (1), all stems with obstruent endings should select *-je*, like *boekje*. If we select all the stems with obstruent endings from the corpus, we can see how many of these actually take the *-je* suffix and if there are any exceptions. This will allow us to calculate thresholds for our hypothetical rules.

Table 2 displays the frequencies of the allomorphs per feature. As we pointed out earlier, it is most desirable if we can have the TP operate directly on CDS. This is why we will use CDS data that was extracted from the Dutch CHILDES corpus (MacWhinney, 2000) available on PaQu. The frequencies were counted as specified in Section 3.2.

---

1. Double forms like *bloempje/bloemetje* 'flower-DIM' are both counted.

**Table 2.** Type frequencies of the subregularities within the CDS corpus

Allomorph	Total	Obs.	Son.	Long	Short	−Pen. Stress	+ Pen. Stress
-je	391	391	0	0	0	0	0
-tje	225	0	225	127	98	1	97
-kje	2	0	2	0	2	0	2
-pje	20	0	20	12	8	0	8
-etje	70	8	62	0	62	62	0
Total	**708**	**399**	**309**	**139**	**170**	**63**	**107**

4.   A Tolerance Principle based analysis of Dutch diminutives

4.1   Finding a productive rule system

We will now systematically show how we can find a productive rule system for the Dutch diminutives with the TP, by highlighting the most important parts of our analysis (for more details, see van Tuijl, 2020). For each potential rule we calculate the threshold with the total number of nouns *(N)* the rule should apply to and check whether this threshold is high enough to tolerate the exceptions.

In order to find the proper set of rules for diminutive formation in Dutch, a child learner has to test such rules and discard or accept them. Let's have a hypothetical child guided by the TP consider different hypotheses regarding the diminutive paradigm. What could be a rule for her to begin with? She will start by looking for a main productive suffix. If only one of the five allomorphs followed a productive rule, all others would be exceptions to this rule. This is possible only if the frequency of the productive allomorph is sufficiently high. The allomorph with the highest frequency is *-je*. So, a rule the child might consider is: 'if diminutive, then select *-je*', where the other allomorphs are treated as exceptions (Figure 1).

Diminutives
|
−je *
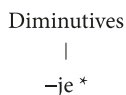
Figure 1. Hypothesis 1

We can determine if this rule is productive by applying the TP to the total number of nouns in Table 2, which is 708. This results in a threshold of $\theta_{708} = 107$ allowed exceptions. This number is far too low to account for the 317 (i.e. 708−391) exceptions in Table 2. As *-je* is the most frequently occurring allomorph, none of

the other allomorphs can hope to be candidates for productive rule formation over the whole system. This shows that the Dutch diminutive paradigm cannot just consist of one productive rule with a number of exceptions. A child learner guided by the TP will not consider this a viable route towards an overall learnable paradigm. As is shown in Figure 1, unacceptable rules are starred *, while acceptable rules will be check-marked ✓ (as will be shown later).

The child who, by the MPP (7), will pursue rules that maximise productivity, will search for an alternative. Instead of settling on a system that has an overwhelming number of exceptions, she will search for other productive rules, "especially rules that divide the data into subsets" (Yang, 2016, p.73).

We will begin our search for subsets with a partition between words with sonorant and obstruent endings (Figure 2). As is shown in Table 2, words with an obstruent ending take -*je* as their suffix almost all of the time. There are only a few cases where a noun ending in an obstruent takes -*etje* as its suffix. These exceptions include words like *trap* ('stairs') and *weg* ('road'), which are diminutivised as *trappetje* and *weggetje* respectively.
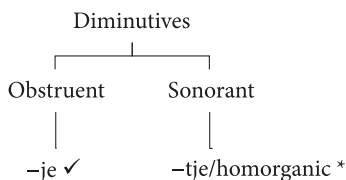

Figure 2. Hypothesis 2

Applying the TP to the subset of words with obstruent endings results in a threshold of $\theta_{399} = 66$. As the actual number of exceptions to the rule within the set of nouns with obstruent endings is 8, this is a perfectly acceptable rule. We can then further test some rules for the set of sonorant ending nouns. We'll start by testing the -*tje* allomorph. The number of exceptions the -*tje* rule can tolerate is $\theta_{309} = 53$. As the number of actual exceptions would be 84, this clearly is not an acceptable rule. However, before completely dismissing this hypothesis, the child might consider a homorganic rule (where -*kje* and -*pje* are derived from -*tje* through place assimilation). On this hypothesis the number of exceptions would be 62 instead of 84. However, as this number still exceeds the threshold of 53, this rule is not acceptable either.

For reasons of space, we cannot present the search illustration in full. For details, see van Tuijl (2020). Instead, we will jump ahead to a system with subsets formed with the features obstruency, preceding vowel length and stress pattern, given the frequencies of the allomorphs within these subsets as shown in Table 2.

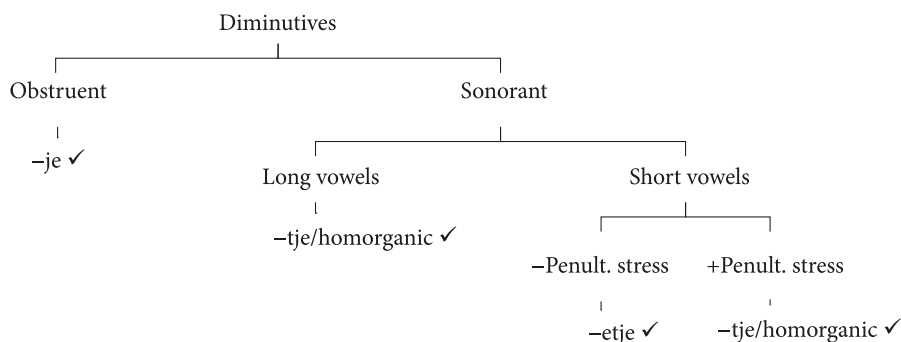The checkmarks in Figure 3 indicate that all the rules in this system are TP-acceptable.

Figure 3. Hypothesis 3

It now appears that we have found a rule system for the Dutch diminutives that accounts for all diminutives in the paradigm and is acceptable under the TP. There is a learning path for the child to follow, determining exactly what works as a productive rule over subsets of the available noun data, and how many exceptions are tolerated. That is what we aimed at in our analysis.

However, there is more than just the corpus of CDS data to base our overall analysis on. As we pointed out earlier, acquisition research has shown that *-etje* is the allomorph that Dutch learning children struggle with the most. As seen in (2) in the descriptive overview of the diminutive domain, three features are needed to describe *-etje* where the others only need one. The *-etje* allomorph is the allomorph that children seem to master last, and it is not only children who have trouble in applying their knowledge of this allomorph consistently. Recall Boersma's (2018) finding that adults often use the *-tje, -pje, -kje* allomorphs where *-etje* is expected, suggesting that *-etje* does not come about as the result of a productive rule, even for adults. So, if the TP indicates that the *-etje* rule can be learned, why does it seem as if neither children nor adults have it as a productive rule? Even though all the rules are TP-learnable, the production facts suggest that the underlying rule system may be structured differently. So let us backtrack to the first subdivision in the data, between obstruent and sonorant endings, and consider some alternative rule options.

Table 3 presents a new distribution of the CDS data. Aside from the three features we have used to find subsets so far, there are some other features we could consider. We have in mind here such features as 'ending in *-m*' and 'ending in *-ŋ*'. This new distribution shows the subsets of words ending in /m/, /ŋ/ and a subset with the rest of the sonorant-ending nouns. This remaining set contains liquidae,

glides, vowels and other nasals. After the split between sonorant and obstruent, the child could hypothesise that there is one rule that states "if a noun ends in *-m*, select *-pje*" and another one "if a noun ends in *-ŋ*, select *-kje*", with both rules as proper instances of homorganic allomorphy (see (3–4) above). Moreover, both rules would be TP- acceptable. Their respective thresholds are $\theta_{28} = 8$, with 8 exceptions and $\theta_7 = 3$, with 2 exceptions.

Table 3. Type frequencies of subregularities within CDS corpus

Allomorph	Total	Obs.	Son.	-ŋ	-m	Son. excluding -ŋ, -m
-je	391	391	0	0	0	0
-tje	225	0	225	0	0	225
-kje	2	0	2	2	0	0 (2–2)
-pje	20	0	20	0	20	0 (20–20)
-etje	70	8	62	5	8	49 (62-5-8)
Total	**708**	**399**	**309**	**7**	**28**	**274 (309 -7 -28)**

In the final column of Table 3 both sets are excluded from the remaining sub-set of nouns with sonorant endings. If we recalculate the acceptability of the *-tje* rule with this data set, the threshold becomes $\theta_{274} = 48$. The number of exceptions is 49. While this number of exceptions is too high, it is, in Yang's (2016, p. 88) wording, "agonisingly close" to being tolerable. See van Tuijl (2020), where the analysis was also applied to a list of the 500 most common nouns in Dutch. If we calculate the *-tje* rule with this distribution, the number of exceptions, 31, remains under the calculated threshold of $\theta_{208} = 38$.

What we would like to propose is that within the set of sonorant-ending nouns, the nouns ending in /m/ or /ŋ/ have their own productive rules. If these latter subsets are excluded from the overall set of sonorant-ending nouns, the remaining *-tje* rule suddenly comes very close to being acceptable as a rule over the sonorant rest set. Figure 4 captures our final hypothesis, compatible with TP-reasoning and the acquisition data available. Under this hypothesis we can also explain the appearance of *-etje* not being the result of productive rule formation. It doesn't need to be productive. Rather, all nouns diminutivised with *-etje* can themselves be treated as exceptions.
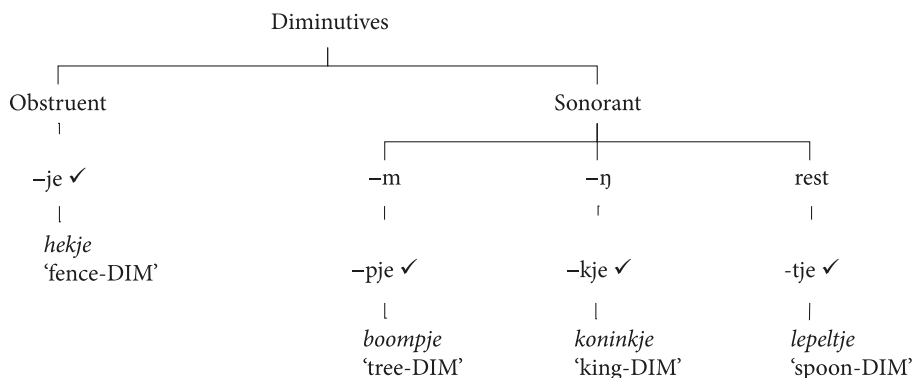
Figure 4. Hypothesis 4

4.2   Discussion

Let's start by mentioning that our reasoning has, of course, been limited by the original morphophonological features we have chosen to consider. In a more thorough analysis other features like the number of syllables might be considered as well. Within our expressly limited view of the paradigm, our analysis shows that a rule system can indeed be found for the Dutch diminutives in which all exceptions are tolerable. However, as described in Section 2.3, complications surface in the acquisition and use of *-etje*. This allomorph appears the most difficult one to acquire, which is not altogether surprising as its descriptive generalisation (2) shows the complexity in combining reference to three phonological features at the same time. Furthermore, it also seems to be unproductive for at least a number of adult Dutch speakers.

Further reasoning has led us to a system that may explain why the *-etje* suffix appears to be unproductive. We also observed this lack of productivity in our own recent wug-test results (van Tuijl, 2020) in line with the earlier acquisition findings. Figure 4 represents a system in which the nouns that receive *-etje* are the exceptions to the system's rules. As it does not include *-etje* as a productive allomorph, it appears – in comparison with Figure 3 – more in line with the findings of the behavioural studies.

We recall Boersma's finding that adults vary in their use of *-etje* (Section 2.3), with a minority using it productively and consistently, and the majority showing inconsistent use, wavering between various allomorphs. This finding may suggest that we are witnessing here the presence of two rule systems, represented by the hypotheses in Figures 3 and 4, with the latter pointing to the majority of speakers, whose acquisition path we have spelled out here.

In Section 2.2 we mentioned some theoretical accounts that have argued for either *-je* or *-tje* as the default allomorph. However, if we follow the strict line of TP-based reasoning, neither can instantiate the default rule. In our analysis, we tested *-je* as a default suffix for the whole set of diminutives. Although *-je* is the suffix with the highest frequency, it is still not high enough to account for all the other allomorphs as exceptions. This automatically means that all allomorphs with lower frequencies certainly cannot approach productivity, if calculated over the whole set of nouns.

We can then use subregularities found in morpho-phonological accounts to find subsets of nouns within the paradigm, building a redistribution on the basis of the phonological features we mentioned at the outset. We can determine the productivity of the allomorphs over these subsets by applying the TP calculation and the reasoning advocated here. This way the TP allows us to shed further light on the language-theoretical proposals via its ability to detect significant linguistic generalisation. The overall question that, of course, remains is which morpho-phonological theory provides the best fit for such a rule system.

It has not been our intention to offer an an alternative morpho-phonological account. Rather, to determine that a system for all the suffixes being productive can be found. And yet given evidence from developmental research, we suggested that we may perhaps do even better. Now we have two different rule systems. Both are TP-learnable. The second system aligns better with the acquisition findings, but the other system could certainly be learned. This may in fact illustrate the possibility of variation between learners. Once again, the next step would be to investigate if further morpho-phonological theorising can shed light on the best fit for either rule system.

This is where we hope to learn further from expert colleagues in phonological theory. One immediately relevant question, raised by one of such reviewers, is the crucial reliance here, and in Yang's theory more generally, on input forms, with predictions on diminutive allomorphy derivable from properties of the stem. How the quantitative calculation method advocated here would work out against a system relying on properties of output form (such as prosodic well-formedness) is something we are aware of but have no explanatory thoughts on.

5. Conclusion

In this study we have described how, on the basis of frequency data on Dutch diminutive suffixes, extracted from both natural language data and specifically CDS data, we can provide a methodical analysis of the Dutch rule system with the TP. This analysis can be used for a further understanding of how the child

learner can figure out the productivity of the Dutch diminutive system. By tying our hypotheses to known acquisition data we have been able to converge on a system of productive rules, and a possible explanation for why one of the five allomorphs, the *-etje* allomorph, behaves differently. This study offers support for the TP and necessitates further inquiry into the theoretical models suggested for diminutive formation.

Funding

Acknowledgements

References

Augustinus, Liesbeth, Ineke Schuurman, Vincent Vandeghinste & Frank Van Eynde. 2014. "GrETEL: Searching for breadcrumbs in texts." Centre for Computational Linguistics KU Leuven. Accessed 2 July 2020, http://nederbooms.ccl.kuleuven.be/documentation/EducationalModule-v4b.pdf

Berko, Jean. 1958. "The child's learning of English morphology." *Word* 14 (2–3): 150–177. https://doi.org/10.1080/00437956.1958.11659661

Boersma, Tiffany. A. 2018. "Variability in the acquisition of allomorphs." PhD diss., University of Amsterdam.

Booij, Geert. 1999. *The phonology of Dutch*. Oxford: Oxford University Press.

Den Os, Els, & Rita Harder. 1987. "De verwerving van de regels voor meervouds- en verkleinwoordsvorming in het Nederlands." *De Nieuwe Taalgids* 80: 240–250.

Gillis, Steven. 1997. "The acquisition of diminutives in Dutch." *Studies in pre- and protomorphology*, edited by W. Dressler. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Huber, Daniel. 2005. "Some notes about the Dutch diminutive suffix." *Government Phonology Workshop* 2(4):23–19.

van der Hulst, Harry. 2008. "The Dutch diminutive." *Lingua* 118.9: 1288–1306. https://doi.org/10.1016/j.lingua.2007.09.012

Kooij, Jan & Marc van Oostendorp. 2003. *Fonologie: uitnodiging tot de klankleer van het Nederlands*. Amsterdam: Amsterdam University Press.

Linke, Kathrin. 2020. "Diminutive allomorphy." *Taalportaal*, 09 June 2021, https://taalportaal
.org/taalportaal/topic/pid/topic-14020545829701958

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk. 3rd edition.*
Mahwah: Lawrence Erlbaum Associates.

Odijk, Jan. 2015. "Linguistic research with PaQu." *Computational Linguistics in The
Netherlands* 5: 3–14.

Trommelen, Mieke. 1984. *The syllable in Dutch.* Berlin: Walter de Gruyter.

van Tuijl, Rosita. 2020. "The productivity of Dutch diminutives." MA diss. Utrecht University.

van de Weijer, Jeroen. 2002. "An Optimality Theoretical analysis of the Dutch diminutive."
Linguistics in the Netherlands 19: 199–209. https://doi.org/10.1075/avt.19.21wei

Yang, Charles. 2016. *The price of linguistic productivity.* Cambridge, MA: The MIT Press.
https://doi.org/10.7551/mitpress/9780262035323.001.0001

Yip, Kenneth, & Gerry Sussman. 1997. "Sparse representations for fast, one-shot learning."
Proceedings of the National Conference on Artificial Intelligence: 521–527.

Address for correspondence

Peter Coopmans
Utrecht Institute of Linguistics OTS
Utrecht University
Trans 10
3512 JK Utrecht
The Netherlands

p.h.a.coopmans@uu.nl

Co-author information

Rosita van Tuijl
Utrecht Institute of Linguistics OTS
Utrecht University

rositavantuijl@gmail.com

Publication history