The role of language-specific phonotactics in the acquisition of onset clusters[*]

Paula Fikkert and Maria João Freitas
University of Nijmegen / University of Lisbon

1.   Introduction

Recently, the acquisition of onset clusters has been much investigated (Goad & Rose 2004, Pater & Barlow 2003, Jongstra 2003, Ohala 1999, Barlow 1997, Freitas 1997, 2003, Fikkert 1994, among others). Most studies focus on onset clusters in isolation, and only seldom learning paths or development over time is discussed. Valuable as these studies are, this study shows that by considering cross-linguistic differences in the acquisition of onsets, development over time, and the language system as a whole, deeper insight is gained into the phonological system and the representations that children are constructing.

It is a well-known fact that children simplify onset clusters for quite some time before they start producing them correctly. Moreover, they do so in a very systematic fashion. A frequently attested simplification strategy is to select the least sonorous element of the target cluster for production. For onset clusters that obey the Sonority Sequencing Principle (Selkirk 1984, among others) — i.e. obstruent-sonorant clusters — the obstruent is realized. For clusters that do not obey this principle — usually, clusters consisting of a (palatal)-alveolar sibilant plus an obstruent — the obstruent is chosen as well. An account that relies on sonority-based onset selection is able to provide a uniform account for both types of clusters. Others have argued that the selection is only indirectly sonority-based: At first, children only realize the head of the onset constituent, but since the optimal head is the least sonorant consonant, both accounts converge for the initial stage.[1,2] Whether directly or indirectly, all accounts mention that an important role is assigned to sonority in onset selection. The data in (1) provide some examples from English (1a), Dutch (1b) and European Portuguese (1c). The patterns are highly similar in the three languages.

(1)	a.	English child language; data from Amahl (Smith 1973)

        [b̥eːt]	'plate'	[b̥aidə]	'spider'

        [d̥eːɹ]	'tray'	[d̥if]	'stiff'

        [ɡ̊ɔt]	'cross'	[ɡ̊ip]	'skipping'

    b.	Dutch child language; data from Robin (Fikkert 1994)

        [bɑt]	brand[3]	'fire'	[pije]	spelen	'to play'

        [tɪkə]	drinken	'to drink'	[tu]	stoel	'chair'

        [kɛin]	klein	'small'	[taːt]	straat	'street'

    c.	European Portuguese child language; data from João Pedro (Freitas 1997)

        [baʃu]	braço	'arm'	[ta]	está[4]	'it is'

        [te]	três	'three'	[paja]	Espanha	'Spain'

        [kawa]	Clara	*name*	[kaʃa]	escada	'stairs'

However, detailed investigation of longitudinal cross-linguistic data show that (a) there is considerable variation in the learning paths towards the adult target form, even for children acquiring the same language, and (b) children acquiring different languages that have the same surface clusters follow different learning paths. The first type of variation can be accounted for by looking at the child's whole system, as will be shown in Section 3.2 (see also Pater & Barlow 2003). If we consider, for example, *sn*-clusters, Eva first realizes the nasal, and at a later stage, the fricative (2a), whereas Robin has exactly the opposite pattern (2b):

(2)	a.	Nasal > Fricative

        [nœyt]	snuit	'snout'	Eva	(1;6.1)

        [zuːpi]	Snoopy	*name*	Eva	(1;9.8)

    b.	Fricative > Nasal

        [fuːpi]	snoepje	'candy'	Robin	(1;10.21)

        [neːw]	sneeuw	'snow'	Robin	(2;1.7)

It turns out that Eva does not have initial fricatives yet, and therefore chooses the nasal first, whereas Robin produces initial nasals consistently correctly about six months later than initial fricatives. For him the fricative is the best option. In other words, the child's segmental phonological system determines the possible realizations of onset clusters.

The second type of variation is the main focus of this paper. We consider children's developmental patterns in Dutch and European Portuguese, two languages from different language families. On the surface they have fairly similar complex onsets. As we saw in (1bc) both Dutch and European Portuguese, henceforth EP, share two types of onset clusters: obstruent-liquid clusters (henceforth CL-clusters) and /s/-obstruent clusters (henceforth sC-clusters), as shown in (3) and (4) for Dutch and EP, respectively.

(3)	Word-initial surface clusters in Dutch

    CL:	pr, br, tr, dr, kr, pl, bl, kl, fr, vr, χr, fl, vl, sl, χl

    sC:	sp(r, l), st(r), sk(r), sχ(r)

(4)   Word-initial surface clusters in EP
          CL: pr, br, tr, dr, kr, gr, pl, bl, tl, kl, gl, fr, vr, fl, ʒl
          sC:[5]   ʃp(r, l), ʃt(r), ʃk(r), ʃf(r)

On the surface the onset clusters of Dutch and EP are very similar. There are a couple of minor differences, though. In the CL-cluster series, Dutch has velar fricatives, which EP lacks. Dutch lacks voiced velar stops, which are present in EP. EP allows 'tl' clusters, but Dutch does not. Dutch has 'vl' and 'sl', EP 'ʒl'. Both languages have 15 different CL-clusters. In the sC-cluster series, the place specification of the initial sibilant differs. However, if we consider the clusters that children attempt the set of clusters is comparable in both languages.

As children acquire lexical phonological representations on the basis of overt output forms, the null hypothesis is that children learning similar surface clusters will show a similar acquisition pattern. This hypothesis will be refuted in this paper, as children acquiring Dutch and EP clearly follow different strategies and acquire a different phonological system on the basis of similar overt input data. We argue that a number of differences in the phonological systems of the two languages give rise to different analyses of the same surface facts. Thus, whereas previous studies have focused on onset clusters in isolation, this paper argues that full insight into the matter can only be gained by considering the language system as a whole.

The paper is organized as follows. In Section 2, we describe the methodology of the study. Section 3 presents the acquisition data, of which a comparative analysis is presented in Section 4. Section 5 summarizes the conclusions.

2.   The methodology

The data come from two longitudinal corpora of spontaneous speech. The Dutch CLPF database (Fikkert 1994, Levelt 1994) contains data of 12 children acquiring Dutch as their first language. The children vary in age between 1;0 and 2;0 at the start of a one-year period of data collection. DAT-recordings were made every other week during play sessions at the child's home. The EP database contains data of 7 children acquiring the Lisbon dialect of EP as their first language. These children were videotaped during play sessions from between 0;10 and 2;0 at the start of a one-year period of data collection. One child was followed for two years (Freitas 1997).

From both sets of data all targets with word-initial clusters were taken and analyzed. We focus on the description of CL-clusters and sC-clusters here, as those are shared by the two languages. For each of the children the developmental patterns in the production of the two cluster types are described. We compared per child and per language which clusters are produced first and how

different clusters are produced in the course of development. The main focus in this study lies on accounting for the similarities and differences in the developmental patterns.

3. The acquisition data

3.1 Order of acquisition of different types of onset clusters

Comparing the acquisition of CL- and sC-clusters, the developmental patterns in (5) and (6) can be observed for Dutch and EP, respectively, where 'S' stands for sibilant, 'P' for plosive, 'L' for liquid, and 'F' for fricative. Most Dutch children acquire CL-clusters before sC-clusters, but a small subset of the children has the opposite order of acquisition.

> (5) Dutch children
>   a. SP   (Robin, Noortje)[6]
>   b. PL > FL > SP   (Other Dutch children)

All EP children have sC-clusters long before CL-clusters, despite the fact that, as in Dutch, the latter type of clusters are much more frequent in the language than the former (Andrade & Viana 1993, Vigário & Falé 1993). About 10% of the intake, i.e. the words that children attempt to produce, contains words with CL-clusters, whereas only 3% contains an sC-cluster.

> (6) EP children
>   SP > PL > FL   all children

Both the Dutch and the EP children acquire PL-clusters before FL-clusters, and similarly, SP is acquired before SF.

There is a striking difference in the order of acquisition of cluster types: EP children acquire sC-clusters much earlier (around age two), than the Dutch children (around 2;6), whereas the opposite is true for CL-clusters. Dutch children start producing these clusters correctly around age two, which is three to six months earlier than EP children. The obvious question is: what leads to these different acquisition paths? Why do EP and most Dutch children differ in the order of acquisition of clusters? Another important question is why Robin and Noortje behave differently from the other Dutch children, even though their input is likely to be similar to that of the Dutch other children? Before we will turn to these questions in Section 4, let us first look at the two types of clusters in more detail.

3.2 CL-clusters

The first stage for both EP and Dutch children is the same: they realize the least

sonorous element of the cluster. The final stage is the same as well. Intermediate stages show considerable variation for both languages. Four different strategies are attested in the Dutch data. Some children produce plosive-glide clusters, as shown in (7a) and (8a). Others produce clusters in which both members share place of articulation, as in (7b) and (8b). A third group produces just the sonorant (6c–7c). Finally, although not very frequent, vowel epenthesis is also attested (7d–8d).

(7)   Developmental pattern for plosive-liquid clusters (Dutch children)

	Stage 1	Stage 2	Stage 3
a.	P	[PG]	PL
b.		[PA]$_{PoA}$	
c.		[L]	
d.		[PvL]	

where [ ] indicates optional stage

(8)   Plosive-liquid clusters
   a.   Plosive > Plosive-Glide > PL
         [kjant]   krant   /krant/   'newspaper'   Catootje   (1;11.9)
         [tje:n]   trein   /trɛin/   'train'   Catootje   (1;11.9)
   b.   Plosive > (Plosive-Approximant)$_{PoA}$ > PL
         [tlatjəs] blaadjes   /blaɟ͡ʝəs/ 'leaves'   Jarmo   (2;1.8)
         [pʊuk]   broek   /bruk/   'trousers'   Jarmo   (2;3.9)
   c.   Plosive > Liquid > PL
         [lok]   klok   /klɔk/   'clock'   Leonie   (1;10.29)
         [liŋkə]   drinken   /drɪŋkə/   'to drink'   Leonie   (1;10.29)
   d.   [pəlaw]   blauw   /blauw/   'blue'   Tom   (1;6.25)
         [kəlɔk]   klok   /klɔk/   'clock'   Tom   (1;6.25)

Clearly, there is a lot of variation in the way children get to the end state. In part, the choice for PG or PL depends on whether a child has acquired liquids or glides first; in part, some children strive for a maximal sonority contrast in the onset (see Fikkert 1994 for more detail).

   EP children, too, have an optional stage in which just the sonorant is realized, as shown in (9a) and (10a). This stage is followed by one in which the cluster is realized with an epenthetic vowel, as in stage 3 in (9a) and (10b).

(9)   Realization and development of initial plosive-liquid clusters (EP)

	Stage 1	Stage 2	Stage 3	Stage 4
a.	P	[L]	P(V)L	PL
b.		Complex segment		

(10)   EP acquisition data

a.	[bɨsiletɐ]	bicicleta	/bisiklɛtɐ/	'bicycle'	Luís (2;2.27)
	[lo]	flor	/floɾ/	'flower'	Luís (1;9.29)
b.	[tɨreʃ]	três	/tɾeʃ/	'three'	Laura(2;2.30
c.	[prajɐ]	praia	/pɾajɐ/	'beach'	Luís (2.2.0)
	[flor]	flor	/floɾ/	'flower'	Laura(2;4.30)

The difference between stage 2 in (9b) and stage 4 in (9a) is not audible. In both cases the cluster is realized. However, the data show a U-shaped development for some children: an early stage in which the cluster is realized at the surface is followed by a subsequent period with epenthesis between the two members of the cluster, before the cluster surfaces correctly again. Laura's data in (11) illustrate this:

(11)   Laura's data

a.	[flor]	flor	/floɾ/	'flower'	Laura	(2;4.30)
b.	[fɨloɫ]	*idem*	/floɾ/	*idem*	Laura	(2;7.16)
c.	[florʃ]	flores	/floɾɨʃ/	*idem* pl.	Laura	(3;0.5)

Freitas (1997, 2003) argued that the first cluster is actually a complex segment, and only the second time that onset cluster appear are they genuine onset clusters. These clusters behave very similarly to the initial onsets in *quarto* /kʷaɾtu/ 'room', etc., which are considered complex segments, not clusters (Andrade & Viana 1993).

Overall, there appears to be a lot of variation: children show different learning paths probably largely due to their different segmental phonologies. No salient differences between the two languages can be discovered.

3.3 sC-clusters

At first, only the second member of sC-clusters is realized in both Dutch and EP. Some Dutch children have an intermediate stage in which the sibilant is realized, as in (13b), before realizing the cluster in an adult-like fashion.

(12)   Realization and development of sC-clusters (Dutch)

Stage 1	Stage 2	Stage 3
C	[s]	sC

(13)   Development of sC-clusters

a.	[tɛp]	step	/stɛp/	'scooter'	Tirza	(1;11.19)
b.	[su]	stoel	/stul/	'chair'	Tirza	(2;0.5)
c.	[spokə]	spoken	/spokə/	'ghosts'	Tirza	(2;5.5)

If we consider the EP data clear differences emerge. EP children do not use the [s] only intermediate strategy. However, epenthesis is frequently attested, both

in front of the cluster and in front of the obstruent, as shown in (14) and (15bc). This is only reported marginally in the Dutch data (Fikkert 1994: 112).

(14)　Realization and development of sC-clusters (EP)

Stage 1	Stage 2	Stage 3
C	VsC	sC
	VC	

(15)　EP acquisition data
a.	[kɛfɨ]	escreve	/ʃkrɛvɨ/	'write_imp'	Marta	(1;8.18)
b.	[ɨʃteɲɐ]	estranha	/ʃtreɲɐ/	'strange'	Marta	(1;10.4)
c.	[ɨkɛf]	escreve	/ʃkrɛvɨ/	'write_imp'	Marta	(1;11.10)
d.	[ʃtɛlɐ]	estrela	/ʃtrelɐ/	'star'	Marta	(2;1.19)

To conclude this section, whereas no clear differences could be found in the learning paths with regard to CL-clusters, those with respect to sC-clusters differ in important ways. EP children often utilize word-initial epenthesis, while this is rare in Dutch.

4.　A comparative analysis of acquisition of sC-clusters in Dutch and EP

There are two major questions that need to be addressed in this section. First, why do EP children acquire sC-clusters much earlier than CL-clusters, despite the fact that the latter type of clusters is far more frequent? Second, why do EP, but not Dutch children, employ the strategy of using epenthesis before sC-clusters? A final question concerns the Dutch data: why does a small subset of the children acquire sC- before CL-clusters?

In the developmental phonologies of EP children there is a striking coincidence: sC-clusters and fricative codas are faithfully realized at the same time, as (16) shows.

(16)
	a.	Coda fricatives					
		[kakɐ]	casca	/kaʃkɐ/	'skin'	Marta	(1;8.18)
		[gɔʃtu]	gosto	/gɔʃtu/	'I like'	Marta	(1;11.10)
	b.	sC-clusters					
		[kɛfi]	escreve	/ʃkrɛvɨ/	'write'_imp	Marta	(1;8.18)
		[ʃtɛlɐ]	estrela	/ʃtrelɐ/	'star'	Marta	(1;11.10)

This suggests that the initial sibilant in sC-clusters is a coda. The fact that a vowel is added to the beginning of the word, allowing the sibilant to surface in coda position points to a similar conclusion. Codas often are acquired earlier than complex onsets.

EP has a lot of external sandhi, giving rise to massive resyllabification in

various contexts (Mateus & Andrade 2000). A special role is assigned to the very frequent verb *estar* 'to be', which surfaces as [tar], [ʃtar], [iʃtar] in the adult language and is targeted frequently by the EP children. This verb often forms one prosodic domain with a preceding word. As many words end in a vowel, the sibilant is often syllabified as the coda of the preceding word. These external sandhi phenomena provide ample evidence for the child that the sibilant in sC-clusters behaves as a coda. Dutch does not have comparable data.

In the acquisition of Dutch, the realization of coda fricatives is very early and precedes the correct realization of sC-clusters by many months. Therefore, there is no reason to assume the sibilant in sC-clusters to be codas. However, Dutch has a far more complex rhyme structure than EP (Fikkert & Freitas 1997). Whereas EP only allows one consonant — [l, ɾ, ʃ, ʒ] — in postvocalic position within the rhyme, Dutch vowels can be followed by two or more consonants, as in *lamp* 'lamp', *plaatst* 'places, 3P.SG.', etc. These words are often analyzed as consisting of a bipositional nucleus (long vowel) followed by a coda, which, in turn, can be followed by coronal obstruents, which form the appendix (Booij 1995). We therefore investigated whether the appearance of sC-clusters in Dutch child language correlates with the acquisition of complex rhyme structures. In (17) the different developmental paths of initial CL- and sC-clusters, and final sonorant-obstruent (-NC)[7] and obstruent-obstruent (-CC) clusters are represented.

(17)   Different developmental orders
a.	CL-	>	-CC				(Leonie, Tom)
b.	CL-	>	-CC	>	-NC		(Jarmo)
c.	CL-	>	-NC, sC-, -CC				(Elke)
d.	-NC	>	-CC	>	sC-		(Robin, Noortje)
e.	-NC	>	-CC	>	CL- >	sC-	(Catootje, Tirza, Eva)

From inspecting (17) only one strong generalization can be made: if a child has sC-clusters, he or she also has clusters in final position. In other words, sC-clusters imply final clusters. What do final clusters and sC-clusters have in common? The answer seems to be that both have an extrasyllabic position: the /s/ in sC-clusters because it does not obey the SSP; the final C in final clusters falls outside the bipositional rhyme. The data do not show a different behavior for the two types of final clusters.

Recall that Robin and Noortje both had sC-clusters before CL-clusters. Both have postvocalic clusters before having any onset clusters. However, Elke, Eva, Catootje and Tirza also have final NC- and CC-clusters, yet they have acquired CL-clusters before sC-clusters. Frequency cannot account for the difference either: both Noortje and Robin attempt many more targets with initial clusters (particularly of the type CL) than final clusters. It seems therefore that the acquisition of CL-clusters is unrelated to that of sC-clusters. However,

the presence of complex postvocalic consonant clusters in the child's system may provide the child evidence for special allowances at word-edges. Knowledge about the existence of extrasyllabic material at right word boundaries may helps the acquisition of an extrasyllabic position at the left word edge, and hence, sC- onset clusters in Dutch.

5.   Conclusions

We have shown in this paper that it is important to consider the language system as a whole to interpret the data, both to explain differences between children acquiring the same language (i.e. the child's own phonological system determines what optimal realizations for clusters are), and between children acquiring different languages. Although EP and Dutch have similar onset clusters on the surface, children do not necessarily show the same learning paths. The phonological system of the language as a whole provides the child cues for analyzing the overt input forms. The analysis has shown that the initial sibilant in sC-clusters is analyzed as a coda in EP. To realize this coda children often produce an initial vowel in early stages of acquisition. This vowel does only surfaces in running speech due to external sandhi in adult EP. For Dutch, the appearance of sC-clusters requires knowledge of extrasyllabicity, and correlates with final clusters.

The differences between Dutch and EP child data can hardly be ascribed to ease of perception and/or articulation, nor to a universal order of development. It also seems that frequency is not playing a significant role either: In both EP and Dutch CL-clusters outrank sC-clusters by far. Yet, some Dutch and all EP children acquire the latter type of cluster earlier. Why some Dutch children acquire sC-clusters before CL-clusters remains a puzzling fact, which warrants further research.

Notes

1. For /s/-sonorant clusters the two accounts make different predictions: a sonority-based account predicts the /s/, a head-based account the sonorant to surface. For an elaborate discussion the reader is referred to Jongstra (2003) or Goad & Rose (2004). Here, we focus on sC- and CL-clusters, as both languages share these.

2. Fikkert (1994) and Goad and Rose (2004) argue that because of differences in sonority profiles children are able to assign different syllabic structure to CL- and sC-clusters. In other words, the SSP is guiding the learning of prosodic structure.

3. Phonetic transcriptions of adult forms:

Dutch: blaadjes /blaȶ͡jəs/, blauw /blauw/, brand /brant/, broek /bruk/, drinken /drɪŋkə/, klein /klɛin/, klok /klɔk/, krant /krant/, spelen /spelə/, stoel /stul/, straat /strat/, trein /trɛin/.

Portuguese: braço /brasu/, três /tres/, clara /klarɐ/, está /(ɨ)ʃta/, Espanha /ʃpɐɲɐ/, escada /ʃkadɐ/.

4. EP spelling suggests that these words start with a vowel, but this vowel is not produced in the Lisbon dialect of EP.

5. In addition to the clusters in (4) EP also has ʒb(r), ʒd(r), ʒg(r), ʒv. Thus, EP has voiced sC-clusters, which Dutch lacks. However, targets with voiced sC-clusters do not occur in the database.

6. Robin and Noortje did not produce any CL-clusters during the recording sessions.

7. Nasal-Obstruent clusters are acquired before Liquid-Obstruent clusters, but -NC stands for both.

References

Andrade, Ernesto & C. Viana (1993). 'Sinérese, diérese e estrutura silábica'. In *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística*. Lisbon: APL. 31–42.

Barlow, Jessica (1997). *A constraint-based account of syllable onsets: Evidence from developing systems.* PhD Dissertation, Indiana University.

Booij, Geert (1995). *The phonology of Dutch.* Oxford: Oxford University Press.

Fikkert, Paula (1994). *On the acquisition of prosodic structure.* PhD Dissertation, HIL, Leiden University.

Fikkert, Paula & Maria João Freitas (1997). 'Acquisition of syllable structure constraints: Evidence from Dutch and Portuguese'. In *Language acquisition: Knowledge representation and processing. Proceedings of GALA '97.* Edinburgh: Edinburgh University Press. 217–222.

Freitas, Maria João (1997). *Aquisição da estrutura silábica do Português Europeu.* PhD Dissertation, University of Lisbon.

Freitas, Maria João (2003). 'The acquisition of Onset clusters in European Portuguese'. *Probus* 15: 27–46.

Goad, Heather & Yvan Rose (2004). 'Input elaboration, head faithfulness, and evidence for representation in the acquisition of left-edge clusters in West Germanic'. In Kager et al. 109–157.

Jongstra, Wenckje (2003). *Variation in reduction strategies in Dutch word-initial consonant clusters.* PhD Dissertation, University of Toronto.

Kager, Rene, Joe Pater & Wim Zonneveld (eds.) (2004), *Constraints in phonological acquisition.* Cambridge: Cambridge University Press.

Levelt, Clara (1994). *On the acquisition of place.* PhD Dissertation, HIL, Leiden University.

Mateus, Maria Helena & Ernesto d'Andreade (2000). *The phonology of Portuguese.* Oxford: OUP.

Ohala, Diane (1999). 'The influence of sonority on children's cluster reductions'. *Journal of Communication Disorders* 32: 397–422.

Pater, Joe & Jessica Barlow (2003). 'Constraint conflict in cluster reduction'. *Journal of Child Language* 30: 487–526.

Selkirk, Elizabeth (1984). 'On the major class features and syllable theory'. In Mark Aronoff & Richard Oehrle (eds.), *Language sound structure*. Cambridge: MIT.

Smith, Neil (1973). *The acquisition of phonology: A case study*. London: CUP.

Vigário, Marina & I. Falé (1993). 'A sílaba do Português fundamental: Uma descrição e algumas considerações de ordem teórica'. In *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística*. Lisbon: APL. 465–478.