1 *Preliminaries*

1.1 Background

I assume that humans have as one module of the mind/brain a faculty of language in the sense of Chomsky (e.g. 2000). The domain of this module is knowledge of language: more specifically of 'I-language' – the individual's internalised knowledge of his or her mother tongue. As an idealisation characteristic of all scientific endeavour, this module, which is a psychological construct, can be treated independently of other cognitive systems and can itself be broken down into a number of sub-systems. That is, the language faculty has internal structure such that it makes sense, for the sake of theoretical investigation, to isolate language from memory, morality and music, and phonology from syntax, morphology and semantics (see Chomsky, *passim*; Hauser *et al.*, 2002; Smith, N.V., 2004; Carruthers, 2006, 2008). On these assumptions phonology constitutes a natural sub-part of the study of I-language, and one area of phonological investigation is devoted to studying the acquisition of phonological knowledge by the child exposed to primary linguistic data from the ambient language.

The traditional formulation has it that this module consists of two components: a lexicon and a computational system (C_{HL} – the Computation for Human Language).The lexicon consists of entries which relate LF (logical form) to PF (phonetic form): more accurately, representations of meaning to representations of sound (or sign). The computational system comprises the syntax and the phonology, which together use these entries to build up paired representations of the meaning and pronunciation of sentences. Except for some discussion of the role of gesture and its position in the grammar, the focus of what follows is restricted to the phonology.

1.2 Phonological theory and phonological acquisition

It is generally accepted that phonological theory and child phonology should and can inform each other (e.g. Menn, 1980; Hayes, 1999; Gierut, 2008), even if

they do so less than is ideal (Fikkert, 2007: 537). This need can be seen clearly in the problem raised by the tension between description and explanation, and the centrality of acquisition in resolving this tension. Linguistic theory, including phonological theory, disposes of a plethora of descriptive devices with overlapping purview – hence potential redundancy. If it can be shown, for instance, that some of these devices are not learnable or give rise to insuperable processing problems, and are implausibly innate, then they should be dispensed with in favour of others that do not suffer from such disadvantages. The problem is pervasive, but can be seen at its starkest in any theory (like that in APh) which makes crucial use of extrinsic rule ordering. Given ten rules, there are 10! (10 factorial, i.e. 3,628,800) ways of ordering them.[1] As Chomsky put it in an early article with George Miller, "we cannot seriously propose that a child learns the values of 10^9 parameters in a childhood lasting only 10^8 seconds" (Miller and Chomsky, 1963: 430).[2] It follows that theories of the acquisition of phonology that have any aspiration to achieving psychological reality must eschew rule ordering. Similar problems arise with theories that make implausible assumptions about any other aspects of learnability: the nature of the child's representations, the limits of variation, the relation between perception and production, etc. (For perceptive general discussion see Velleman and Vihman, 2007.)

1.3 Perception and production

Although the isolation of phonology from other systems is essential, some aspects of phonological acquisition can be explained by reference to properties external to the developing child's strictly linguistic abilities – specifically, perceptual and motor maturity. There is a universal asymmetry in perception and production such that children can demonstrably perceive contrasts that they do not, often cannot, produce (for an overview of the infant's perceptual abilities, see Eimas, 1996; Jusczyk, 1997). In reaffirming this truism it is important to stress that 'perception' is not a simple notion. At a few weeks of age infants are sensitive to the statistical properties of the input language and are able to extract relevant information from it. Thus Saffran *et al*. (1996) demonstrated that 8-month olds could identify 'words' consisting of sequences of three syllables purely on the basis of the relative frequency of such sequences.

1 There are ways of decreasing the numbers involved but, as I postulated twenty-nine ordered rules, the magnitude of the problem should be clear.

2 The notion of parameter appealed to is not the same as the current notion discussed below. The length of childhood is perhaps a little brief: 10^8 seconds is a little over 3 years.

Subsequent research (as reported in e.g. Saffran, 2001) has shown that this ability generalises to non-linguistic domains, most notably music, and presumably reflects a domain-general cognitive ability.

To a first approximation it is fair to say that 6-month olds can cope with anything that the world's languages can throw at them, showing categorical perception of contrasts in, for instance, Voice Onset Time at a few weeks of age. However, it is striking that infants at 6–8 months perceive phonetic contrasts better than they do at 10–12 months and often better than adults (Werker and Stager, 2000: 183; for a summary, see Pater, 2004). In fact, such a developmental progression is characteristic of infant abilities more generally. At 3 months children have sensitivity to a greater range of faces (e.g. Caucasian, African and Chinese) than at 6 or 9 months (Kelly *et al*., 2007; Pascalis *et al*., 2005), a period during which their abilities for recognising 'own-race' faces improves.[3] The parallel with language is strikingly close. Further, perceptual sensitivity seems to be dependent on what the infant is doing. An ability which is demonstrably deployed in simple discrimination tasks (differentiating [ba] and [pa], for instance), may not be used in a word-learning task: "Infants are listening for different information in phonetic tasks as compared to phonological tasks" (Werker and Stager, 2000: 190). Even children at 20–6 months often confuse minimally distinct words unless they know them well (Barton, 1976): that is, familiarity based on frequency is one relevant determinant of 'perception'. None the less, it seems to be the case, as Mani and Plunkett (2007: 252) put it, that "part of the infant's phonological repertoire appears to be in place before lexical acquisition is set in motion". How much of that repertoire is established is contentious and probably varies from child to child. Pater (2004: 223) argues that at 14 months "the consonantal place distinction is not encoded in lexical representations, though it is present in phonetic representations". A further complication is provided by perceptual magnet effects (Kuhl, 1991) – a kind of shrinking of the perceptual space so that there is poor discrimination near phonetic category prototypes of the native language. This is a further manifestation of categorical perception (cf. e.g. Feldman and Griffiths, 2007) but one where the role of the native language is paramount.

The relation between perception and production has various implications for any theory of the acquisition of phonology. Does the asymmetry affect the child's knowledge (competence) or only his use of that knowledge in performance? What is the nature of the child's lexical representations? How many levels

3 In face recognition the sensitivity generalises to other species, especially monkeys, suggesting that it is a property of primates (or mammals) rather than just humans.

of representation is it necessary or desirable to postulate? In particular, is there a single lexicon or should one postulate a dual lexicon? Is the Optimality-Theoretic resolution of the "comprehension/production dilemma" (Smolensky, 1996) convincing? In APh I claimed – incorrectly as it transpired – that the child's perceptual abilities were essentially flawless. I return to this in detail in chapter 2; for the moment we turn to the first issue listed: the traditional (Chomsky, 1965) distinction between competence and performance.

1.4 Competence and performance

The first issue to be addressed is whether the acquisition of phonology should fall under a theory of competence or a theory of performance, or should be compartmentalised judiciously between the two. Pronunciation is relational in nature: it relates abstract mental representations to articulatory and auditory sequences which have acoustic properties; that is, the child's representation(s) must be or become "legible" (Chomsky, 2002) at the motor and cognitive interfaces. In the case of the establishment of pairings between representations of sound and representations of meaning – that /ˈpeŋgwɪn/ conveys the meaning PENGUIN – it is clear that we are dealing with knowledge of language in the traditional Chomskyan sense, so we are investigating the learning child's competence as we study this aspect of his developing abilities. Matters are less transparent when we study the child's pronunciation of *penguin* as he acquires his phonology. If (like Z at one stage) the child pronounces it [ˈb̥ɛmiː] this could be because of his different competence – his representation, for whatever reason, is different from that of the adults around him – or it could be that his competence (in so far as it pertains to the phonology of the lexical representation) is the same as the adults' and the pronunciation is a matter of performance. A combination of these two options is also plausible: the child's representation may be partly correct and partly incorrect. For example, it might be both significantly underspecified and also subject to distortion in performance. It also bears mentioning that the effect of the ambient language on the child's babbling may be important. Boysson-Bardies (1999) shows that the phonetic properties of the child's babbling are in part a function of the sound system he is exposed to before elements of his vocabulary have been established, indicating that cognitive representations are not necessary concomitants of pronunciation.

The allocation of responsibilities to competence or performance relates to at least two further issues. The first is whether the child 'has his own system'. Are his deviations from adult pronunciation a function of his manipulating an idiosyncratic phonology or simply a distorted reflection of the adult system

being acquired? This amounts to asking whether the 'realisation rules' attributed to the child (see below) are a matter of 'incompetence'or 'malperformance' (APh: 140; cf. the discussion in Hale and Reiss, 2008). The second pertains to the role of UG (the "theory of the initial state" – Chomsky, 2000: 81) in the child's arrival at whatever state of knowledge underlies his performance. To what extent (if any) is the child's developing phonology determined by properties peculiar to language, and to what extent is it determined by properties which are system-wide? To take a specific example: are the perceptual abilities mentioned above peculiar to language or (more plausibly for those of them which are shared with chinchillas – Kuhl and Miller, 1975) general to the whole of audition? On the production side, the child's nascent imitative abilities might be at least in part the result of the action of mirror neurons (see e.g. Iacoboni, 2008) which are not restricted to language, or they could be tied to the linguistic domain. There is intriguing evidence from the acquisition of sign language that there is an "innate agenda" (Petitto, 2005: 95; cf. Morgan, 2006) which is neutral between the modalities of speech and sign, but is specific to language in that it has no known reflexes elsewhere in the organism. This early language-specificity is corroborated by later dissociation of pointing and signing – both in sign language acquisition where there are rare cases of pronoun reversal (Meier, 2002; cf. Chiat 1986), and in sign language loss in aphasia where, depending on the site of the lesion, the same physical movement may be retained as a gesture and lost as a sign or vice versa (Poizner *et al*., 1987).

The ultimate aim is to explain the child's behaviour and this will necessitate some consideration of both competence and performance. Their interplay is rarely transparent but, to anticipate the discussion to come, I think that (virtually) all non-adult pronunciations are a function of performance: 'malperformance' in the terms used above. This is despite the systematicity of the child's production and despite the fact that there may be grammatical (competence) determinants of phonological patterns. For example, the development of final /z/ in A's production was determined in part by whether it corresponded to the plural morpheme, as in *peas*, or to one segment in a mono-morphemic word, as in *please*. Importantly, just because the operation of the child's production system is partially determined by considerations of competence it doesn't follow that the processes are themselves part of the child's competence.

If the child's output is a matter of (mal)performance it can be characterised by a neural network which associates a phonological representation as input with articulation as output. Because the child's output is systematic, deviations from the adult forms can be described in terms of 'rules' which have the appearance of constituting a competence grammar. This appearance is reinforced by the

effects of grammatical knowledge, so that the input to the postulated neural network cannot simply be an adult 'phonemic' string, but must be a grammatically parsed string including such information as {plural}, 'belongs to a functional category' (see the next section), and so on. If correct, this position has significant implications for the nature of the acquisition of phonology, in particular the levels of representation it is necessary to postulate.

1.5 Levels of adequacy

Any theory should meet certain 'levels of adequacy' (Chomsky, 1964, 2004, 2009a; see also Smith, N.V., 1989: ch.11; 2004: 58f.): observational, descriptive and explanatory,[4] where the last of these is standardly taken to be equivalent to 'acquisitional adequacy', the property of a theory that accommodates the possibility of first language acquisition. Ideally the theory should also attempt to go 'beyond explanatory adequacy' (Chomsky, 2004), deriving specific properties of language, child phonology in the present case, from outside the phonology (e.g. phonological processes which are epiphenomena of grammatical ones) or even from outside the language faculty. An example of the former, where phonological phenomena are a reflex of grammatical facts, was illustrated above by the appearance of final [z] in A's developing language. In adult English, final [z] may be either the last segment of a monomorphemic word (e.g. *cheese*), or any one of four different morphemes: plural (as in *eyes*), third person singular (as in *throws*), the reduced form of *is* (as in *Daddy's going*) and the possessive (as in *Daddy's one* or *It's Daddy's*). The fact that these different examples were all treated developmentally differently by A (see APh: 67ff.) suggests strongly that non-phonological factors need to be taken into account. A different kind of example showing the relevance of morphological structure in accounting for phonological development comes from A's treatment of unstressed initial syllables, which were systematically replaced by [riː]: e.g. *attack* → [riː'tæk] (APh: 172f.).

Similarly, in Z's development, initial /ð/ was systematically omitted (unlike initial /θ/ or non-initial /ð/) giving rise to pronunciations such as [ɛn] for *then* and [is] for *this*. When I invented neologisms such as *thub* ([ðʌb]) beginning with /ð/ to represent new toys, Z consistently produced them with initial [z], so

4 Although these levels represent increasing degrees of success, progress is not necessarily linear. Hayes (1999: 247) makes the nice point that we often have phonetic explanations for phonological facts even when we are unable to incorporate these explanations into a descriptively adequate formal phonology.

his omission of all and only initial examples of /ð/ is presumably due to the fact that all /ð/-initial words in English belong to 'functional categories'. A parallel example is provided by his idiosyncratic distinction between *you*, *your*, *yours*, *yourself* on the one hand and all other words beginning with /j/ such as *yap*, *young*, *yoyo*, etc. on the other. All the former lost the initial consonant entirely, giving [uː, ɔː, ɔːz, ɔːˈsɛlf]; all the latter replaced the /j/ by [l], giving [læp, lʌn, ˈləuləu], etc.

Examples of data from phonological acquisition which can be explained by appealing to facts outside the language faculty are manifold, even if frequently contentious. The most obvious examples are phenomena which are explicable on the basis of the child's perceptual immaturity or motor inability. The widespread pronunciation of words like *bottle* and *puddle* as [bɔkəl] and [pʌgəl] may well be due to a failure to perceive the difference between [d] and [g] before a 'dark' [ɬ], though we shall see below that such an explanation is only partial. Similarly, simple articulatory inability may be sufficient to explain why children typically go through a stage in which all of *mash*, *mat*, *mass* and *match* are pronounced identically as [mæt]: the child may be incapable of producing the fricatives and affricates which distinguish the items in the adult language. A second example can be taken from variation in fine phonetic detail: different degrees of nasalisation of vowels adjacent to nasal consonants may have no phonological significance but simply reflect motor development constrained by a universal physiological principle of least effort.

It is in principle not difficult to test for the role of such 'external' factors, even if some of the further subtleties which need to be teased out are a matter of dispute. However, there is a clear difference between such examples and 'chainshifts' of the kind seen in 'puzzles' or metathesis. A pronounced *puddle* as [pʌgəl] but *puzzle* as [pʌdəl], indicating that motor inability *tout court* was implausible as an explanation for the former mispronunciation. Similarly, as the result of a regular process of metathesis, he pronounced *icicle* as [aikitəl] though he produced the invented word '*aikitəl*' as [aikikəl]. The same conclusion that 'motor inability' is only a partial explanation for the child's productions can be derived from cases of free variation between a correct and an incorrect form. For instance, Z's pronunciation of *rain* as [rein] or [wein] (while *wet* appeared only as [wɛt] and never as [rɛt]) shows that something more than articulatory incompetence must be at stake. It may of course be that in this case what is lacking is precisely the ability to control the production of [r] consistently but such an account does not fully generalise, as there are further examples of mispronunciations which are not obviously due to production difficulty and are implausibly attributable to perceptual confusion.

The first of these is recidivism, a diachronic variant of the free variation just mentioned. In this case the child produces the correct adult pronunciation at one stage and then regresses to an incorrect form: for instance, Z pronounced *red* correctly as [rɛd] for some four months before adopting the pronunciation [wɛd]. It is not impossible that motor control should be lost but it is not the most plausible hypothesis. The second kind of example is also a special case of the free variation mentioned in the previous paragraph, illustrated in Z's case by an asymmetry in his treatment of Cl/Cr clusters. The sequences [br] and [pr] were used for both /br, bl/ and /pr, pl/ respectively, but [bl] and [pl] were never used for /br/ or /pr/. Given a pattern where /bl/ is pronounced as [br] or [bl], but /br/ is pronounced only as [br], it is not plausible to attribute the pronunciation of *blood* as [brʌd] merely to motor inability to produce [bl]. More interestingly, this example demonstrates that the child must distinguish the clusters 'br' and 'bl' in his lexical representations though he may not be able to control the production of one of them ([bl]) adequately.

A third class of examples is provided by variation where the child's pronunciation is determined by properties of the adult form unpronounced by the child: for example, Z pronounced *all* either with or without the final /l/, as in [ɔː ə ləud], *all the load*, versus [ɔːl in ɛː], *all in there*. The variation was not random; rather, /l/ was omitted before an adult consonant but retained before an adult vowel, even though the relevant adult consonant (/ð/ in this case) was not itself pronounced. Further examples, like the contrast between ['pɔːriʔ] (*pour it*) and ['pɔː ə tiː] (*pour the tea*), suggest that the phenomenon was more general, though in this case the presence of linking [r] in the input makes interpretation more difficult.

Scobbie (2007: 21) suggests that the child's mental representations could be "non-deterministic", but the claim seems implausible in the light of this evidence. It may be that there is a stage when certain (e.g. parametric) decisions have not yet been made, and may perhaps never be made (see Smith, N. V. and Cormack, 2002), but the lexical representations must be determinate to allow a coherent account of the range of data cited above. Take, for instance, the final example of *all the load*: a natural assumption might be that the initial segment corresponding to adult /ð/ is underspecified, hence indeterminate – it could be that there is just C there. Despite its superficial plausibility such an analysis would predict other errors: that /ð/ should on occasion be replaced by some other consonant and not just deleted, but such errors were not attested.

1.6 Levels of representation and the units of representation

These examples raise complex issues about what precisely the child acquiring his first language is representing. To investigate this we need as a preliminary to

specify what the child is taken to be acquiring when he 'acquires phonology'. I assume that he must learn the phonological representation of the lexical items of the ambient language in terms of 'phoneme-sized' units decomposed into distinctive features. These representations are neutral as between perception and production and so the child also has to learn to relate them to auditory percepts and to sets of articulatory instructions. These instructions need to specify the fine phonetic detail characteristic of the adult language. This includes such subtleties as the distinction between clear [l] and dark [ɫ] in English or between alveolar [t] in English and dental [t̪] in French, even though these differences are not consciously accessible. That they are none the less part of the speaker's tacit knowledge is evident from our sensitivity to individual differences or to foreign accents, where these are characterised precisely in terms of such sub-doxastic properties. The child also has to learn the patterns of stress and intonation of the target language and how to modulate the phonological and phonetic properties of lexical representations in syntactic context. About these I shall have little to say, though they do raise the question of the relation between the phonetic and the phonological and the extent to which it makes sense to draw such a distinction in the adult language. On this latter point there is little consensus. Whatever decision is taken in this regard, we then need to investigate whether the same levels are plausible for characterising the child's nascent phonology or whether he disposes of more or fewer levels than the adult. Here there is even less agreement. Comparable questions then arise with regard to the units of representation it is necessary to postulate: distinctive features, elements, phonemes, and so on.

Along with the majority of phonologists I propose simply to accept the validity of the phonetics–phonology distinction without much further discussion, though evidence will be presented concerning the number of levels of representation it is necessary to postulate. On one interpretation of this evidence, the phonetics–phonology distinction is moot (for interesting discussion see Hale and Reiss, 2008: ch. 6). For present purposes, I assume the classical distinction between lexical and 'surface' levels, a distinction reaffirmed more recently by Boersma (2006: 1) who writes that "the minimum number of representations that we need to do interesting phonology [is] two phonological representations connected to each other and to two semantic and two phonetic representations". The phonological representations are underlying and surface; the phonetic representations are auditory and articulatory. This still doesn't exhaust the issue of what kinds of representation the infant speaker-hearer has to manipulate and whether these change over time. For instance, adults might deploy a system with at least phonological and phonetic representations,

whereas the child might go through a stage which is 'pre-phonological' (or at least pre-segmental – cf. the discussion of recidivism in section 5.1.3) at which any representation is an undifferentiated gestalt, or purely phonetic without phonological structure. Another possibility, discussed in section 5.1.2 below, is that the child's own pronunciation is not represented at all, in which case the articulatory representation becomes superfluous.

Whatever the decision on that issue, the assumption that we need both auditory and articulatory representations raises the problem of the need to define the distinctive features whose combinations enter into both phonological and phonetic representations. Here there are at least four possibilities: that the features be defined articulatorily, acoustically, auditorily or abstractly. For some the auditory percept of an utterance is crucial, for others its acoustic properties. Kingston (2007) provides a useful overview, concluding (p. 432) that the evidence suggests that what is important is "the auditory effects of the signal's acoustic properties". Given the need to provide a processing basis for both hearer and (adult) speaker, as well as a means for lexical storage accessible to both, some amalgam of the kind Boersma presupposes is perhaps plausible. (See Harris, 2007: 124ff., for discussion.)

For both adult and child I assume that we need distinctive features of the general kind presented in Chomsky and Halle (1968, hereafter 'SPE') as updated and presented in Hall (2007), and that these are at the base of a phonological hierarchy. This consists of at least phonological words, which consist of feet, which consist of sequences of syllables, in turn broken down into onsets and rhymes (consisting of an obligatory nucleus and an optional coda), and where all of these constitute 'phoneme'-sized segments composed of said distinctive features. I am agnostic about the need for a constituent 'rhyme',[5] but Z's phonological development provides evidence for the necessity of postulating onsets as constituents. In my APh (pp. 170, 188f., 190, 191) I argued for (and against) distinctive features, and for segments and syllables.

The phonological hierarchy will also need to interface with the syntax in order to accommodate the kind of morphological effects (plurals, the *attack* examples) discussed above. Words, and perhaps higher units both phonological, such as 'intonational phrases' as in Truckenbrodt (2007: 436), and syntactic (phrasal projections), are presumably necessary but I have nothing useful to say about them.

5 I suspect that it is necessary, but A's and Z's phonological development provided no direct evidence.