

PHONOSEMANTIC SUBSETS IN THE LEXICON: HUNGARIAN AVIAN NOMENCLATURE AND *L'ARBITRAIRE DU SIGNE*

DANIEL ABONDOLO

University College London, School of Slavonic and East European Studies

BEGINNING with the premise that complete arbitrariness of the linguistic sign is unlikely if not impossible,¹ I examine in this essay non-derived Hungarian avian nomenclature in an attempt to assess degrees of arbitrariness (and its opposites) in the fit of form to function.² Two fundamentally different, indeed opposed, kinds of non-arbitrariness³ are distinguished: self-similarity among items in semantic subsets and, at the same time and in the same subsets, strongly disparate deployments of phonemes and shapes of roots. Self-similarity arises from two sources, again fundamentally opposed: divergence and convergence. Disparate phoneme deployment and root shape also have two opposite sources: old versus relatively new lexis. Various kinds of 'iconicity' are distinguished, and iconicity as a whole, narrowly defined, is distinguished from indexicality, which of these two kinds of non-arbitrariness is by far the more prevalent. At least in the subset of bird names, indexicality seems to reside in clusters of self-similar forms. Within and across these clusters, iconic features of various kinds enrich the picture by encoding relative size of the referent. Near the end of the essay I put forward an etymology for *madár*, 'bird'.

¹ See, for example, Benjamin K. Bergen, 'The Psychological Reality of Phonaesthemes', *Language*, 80, June 2004, 2, pp. 290–311. Notice in particular (p. 293): 'On the basis of the assumption that a simple word's phonological form is entirely arbitrary, given its semantics, there should be the same portion of *gl-* words that have meanings related to "light" or "vision" as there are *sn-* words that share those meanings. And yet [...] this is clearly not the case.' Note that the term *phonaestheme* applies to a range of sound-sense pairings that is broader even than the one examined here (see Bergen, p. 290 for literature), but unfortunately the term itself implies a rather narrow compass, the elements *phon-* and *-aesth-* suggesting that only sound and sensation are involved.

² This is the place to express my thanks to Richard Brown and Szilvia Molnar, with whom I had critically helpful conversations while writing this essay.

³ For a recent overview of non-arbitrariness and related matters, see John E. Joseph, *Limiting the Arbitrary: Linguistic Naturalism and its Opposites in Plato's Cratylus and Modern Theories of Language*, Amsterdam, 2000 (hereafter *Limiting the Arbitrary*), especially chapter 4, 'Natural Grammar and Conventional Words from Aristotle to Pinker', pp. 93–140.

I am interested here not in the etymology of the Hungarian names⁴ of the words for birds *per se*, but rather in the present-day shape of the names themselves and the relations these shapes have with those of other names, specifically of Hungarian words for birds and of other Hungarian words in general. To put it another way, I am here not primarily interested in looking to history for explanations, as is, say, Jenő Kiss in his monograph on Hungarian avian nomenclature.⁵

The history of these names (the changes undergone by their forms, and the various ways in which they have been used, that is, the senses with which they have been associated over time) is of interest, but in a synchronic inquiry cannot serve as an explanatory principle.⁶ Rather, what I am suggesting here is the reverse: that synchronic relations between sound and sense within reasonably clearly delimited semantic subsections of the lexicon can provide assistance in our understanding of that history.

An example should make this clear. In the most recent compendious reference-work on the history of Hungarian words⁷ we read that the Hungarian word for 'salmon', *lazac*, is a loanword of Slavonic origin. So much is clear. Nevertheless an enigma surrounds the phonological development of this word. The discrepancy between the phonologies of the Slavonic words cited, all of which have various voiceless sibilants or shibilants (for example, Slovak *losos*, cited as the most probable direct source) and that of Hungarian *lazac*, with voiced fricative *-z-* and voiceless affricate *-c*, is glossed over by the *EWU* as follows: 'Form *lazac* entstand durch Stimmhaftwerden und Affrikation.' As explanation, this is no better than *le malade imaginaire's* explanation that opium is soporific because it contains a 'virtus dormitiva'. Naming is no substitute for writing a history. In fact, as we shall see below, I suggest that the change from **sVs* to *zVc* brought this word more into line with the lexical canon⁸ of Hungarian: it came to sound more like a Hungarian word.

⁴ To avoid perilous misunderstandings, in this essay I adopt the Gombocz/Ullmann system of talking about the words of speech-mediated language as having both names (audible, pronounceable representations) and senses (mental representations, which may be of things or of other representations). Thus the four letters <|>, <0>, <v>, <e> written or printed in that order (<love>) form a visual representation of the (pronounceable, audible) name of a word whose sense is the mental representation of a mental representation. See Stephen Ullmann, *The Principles of Semantics*, 2nd edn, Glasgow, 1959 (hereafter *Principles*), pp. 65–73; and Zoltán Gombocz, *Jelentés és nyelvtörténet* [1926], Budapest, 1997, pp. 147–49. Non-arbitrariness in the names of words in signing is an entirely separate issue.

⁵ Jenő Kiss, *Magyar madárnevek: az európai madarak elnevezései*, Budapest, 1984 (hereafter *Madárnevek*).

⁶ A detailed investigation into the histories of these words would entail meticulous tracing not only of sound changes but also of sense development or even radical shift. For example, on the basis of the phonological record the premier Hungarian etymological dictionary (Loránd Benkő [ed.], *A magyar nyelv történeti-etimológiai szótára*, Budapest, 1976 [hereafter *TESz*]) reconstructs *hőrsög* as having been used to refer to toads (c. 1395), hedgehogs (1522), badgers (1533), marmots (1604), and hamsters (1608); only the last meaning is current. Heterogeneous though these glosses may seem at first glance, one sees that a thread of semantic continuity is maintained: apart from the oldest gloss, taken from a wordlist, not a running text (the *Besztercei szójegyzék* [Glossary of Beszterce], from c. 1390–1410), all the putative referents are mammals.

⁷ Loránd Benkő et al. (eds), *Etymologisches Wörterbuch des Ungarischen*, Budapest, 1992–1993 (hereafter *EWU*), p. 877.

⁸ While my use of the term *canon* implies cleavage to, and deviation from, statistical norms, I attempt no elaborate statistical analysis in this essay. *Statistical norm* here refers to 'a crude intuitive feeling for quantity based on crude counts', see Robert Austerlitz, 'The Morphology and Phonology of Finnish Given Names', in Irmengard Rauch and Gerald F. Carr (eds), *Linguistic Method: Essays in Honor of Herbert Penzl*, The Hague, 1979 (hereafter 'Finnish'), pp. 299–306 (299). For an example of an analysis of the lexical canon of one language, see Paul Menzerath, *Die Architektonik des deutschen Wortschatzes*, 1954 (hereafter *Architektonik*).

So much for the historical background of the names and senses of these words. But I must stress here that the nature of the referents, that is, of the birds themselves, is also peripheral to this study. The qualities of the referents are of interest and importance, as are the cultural values derived from and attributed to them. But the focus here is on the forms of the names themselves. Groupings and patterns of such names may then be compared with groupings in the world (ethnoscience, Western science; technology of everyday life).⁹

Much of my method takes its tools from the workshop of internal reconstruction. The study of a linguistic isolate requires internal reconstruction: the close examination of the lexicon and morphophonemics of a language, and the study of its 'irregularities' — degrees of productivity, doublets and other paronyms, synonymic clustering. But internal reconstruction may also be applied with profit to a language with clear congeners. For example, the juxtaposition of Hungarian *harkály* 'woodpecker' with its ObUgrian pendants is a necessary step in the understanding of the history of this word and of its cultural antecedents and connexions; but my point is that this step is not sufficient. The Hungarian word *harkály* should be studied for itself, and not merely from 'without' and from 'above', that is, comparatively and diachronically. It should also be studied from 'within' and from 'below': it should be examined in the light of the entirety of our knowledge of Hungarian avian nomenclature and in the light of the relations, or valences, of the phonemes and phoneme combinations of which the name of *harkály* is made.

For the purposes of this essay I take a moderately realist position: I assume that birds are a natural kind, and that the names we use to refer to them are thus natural kind terms. The fact that this or that name may refer to different subspecies, species, or parts of different families in no way undermines what I take to be an underlying nature that birds have 'in a way that artefacts don't'.¹⁰ Put another way, 'out-thereness' is not only the consequence of scientific work and everyday thinking but also their (partial) cause.¹¹

On the other hand, I cannot agree with Goddard when he asserts 'if we heard of a scientist gathering some specimens of *chairs* in order to discover the nature of *chairs*, we would think that he or she is crazy',¹² unless we were to move the italicization from the word *chair* to the word *scientist*. For the investigator in the humanities does precisely this sort of work, and is far from 'crazy'; he or she does indeed collect specimens and other data concerning artefacts in order to learn more about them. As William A. Foley puts it:

'Furniture' is not some notion understood simply through a distributed neural network in the brain; it is something our bodies use, sit on, or sleep in. Knowledge is perceptually guided action in this view, bringing forth a meaningful world in which continued structural coupling is possible. This coupling is not optimal; as with evolutionary change, it is simply viable. Knowledge is a kind of lived history.¹³

⁹ Equally of interest are the questions of birdnames cross-linguistically; but these can be addressed adequately only once detailed work within individual languages has been carried out.

¹⁰ Cliff Goddard, *Semantic Analysis, A Practical Introduction*, Oxford, 1988, p. 251.

¹¹ John Searle, *Mind, Language, and Society*, London, 1999, p. 24, citing Bruno Latour and Steve Woolgar, *Laboratory Life: The Construction of Scientific Facts*, 2nd edn, Princeton, 1986, pp. 180–82.

¹² Goddard, *Semantic Analysis*, p. 251.

¹³ William A. Foley, *Anthropological Linguistics: An Introduction*, Oxford, 1997, p. 90. Note also, in connexion with the position taken here concerning name, sense, and mental representations that 'love' is in this way similar to 'furniture' (see note 4).

Whenever we deal with large amounts of data we are faced with the question: are strict quantitative methods in order? As I hinted above, by and large, given the vastness of the lexicon and the kind of work essayed here, I think the answer is no.¹⁴ Intriguing though they may seem at first sight, in our present state of ignorance attempts at setting up a metric of any strictness by which one might measure the degree of prototypicality of a given bird designation are probably doomed to dubiety if not outright failure.

For example, we might set up a scheme with which to evaluate the formal features of names within various semantic fields. The procedure involves positing clusters of shared features.¹⁵ Thus we might specify as core features of Hungarian bird names the presence of s(h)ibilants (*sísegő*)¹⁶ and liquids and, at the morphological level, bisyllabicity and the absence of stem alternations:

	<i>kócsag</i>	<i>sas</i>	<i>gólya</i>	<i>szarka</i>	<i>varjú</i>	<i>galamb</i>	<i>daru</i>	<i>ölyv</i>
gloss	egret	eagle	stork	magpie	crow	dove	crane	buzzard
s(h)ibilant	+	+	–	+	–	–	–	–
liquid	–	–	+	+	+	+	+	(+)
non-alternating ¹⁷	+	+	+	+	–	+	(–)	–
bisyllabic	+	–	+	+	+	+	+	–
'score'	3	2	3	4	2	3	2	1

In this small set, *szarka* emerges as a comparatively prototypical bird name, while *ölyv* seems to be an outlier. If we apply this scheme of values to body-part terms, they score rather poorly:

	<i>szem</i>	<i>száj</i>	<i>szív</i>	<i>fej</i>	<i>kéz</i>	<i>nyelv</i>	<i>ujj</i>	<i>kar</i>
gloss	eye	mouth	heart	head	hand	tongue	finger	arm
s(h)ibilant	+	+	+	–	+	–	–	–
liquid	–	–	–	–	–	+	–	+
non-alternating	+	–	–	–	–	–	–	+
bisyllabic	–	–	–	–	–	–	–	–
'score'	2	1	1	0	1	1	0	2

On the other hand shared features of body-part names¹⁸ seem to be sesquisyllabicity and the presence of 'glides' (*j v*) and liquids (*l r*):

¹⁴ This is not to discount the value of precise, statistically aware work with closed corpora. But the project here is exploratory and hermeneutic: the objective is to discover what we need to know more about.

¹⁵ See John R. Taylor, *Linguistic Categorization: Prototypes in Linguistic Theory*, Oxford, 1989 (hereafter Taylor, *Categorization*), p. 48.

¹⁶ I am grateful to Endre Táló (personal communication) for this Hungarian term.

¹⁷ The term *non-alternating* is used here to refer to the subset of Hungarian stems which undergo neither epenthesis nor alternations of final or penultimate non-high vowels with zero. See D. Abondolo, *Hungarian Inflectional Morphology*, Budapest, 1988 (hereafter Abondolo, *HIM*), p. 180 and pp. 203–04, 209–10, where they are called 'stable' stems, and, for a radically different approach with similar results, András Kornai, *On Hungarian Morphology*, Budapest, 1994, pp. 30–35 and 115–20.

¹⁸ I return to Hungarian body-part terms below.

	<i>szem</i>	<i>száj</i>	<i>szív</i>	<i>fej</i>	<i>kéz</i>	<i>nyelv</i>	<i>ujj</i>	<i>kar</i>
sesquisyllabic	–	+	+	+	+	+	+	–
glide (j v)	–	+	+	+	–	+	+	–
liquid	–	–	–	–	–	+	–	+
no kB ¹⁹ –	+	+	+	+	+	+	+	–
‘score’ ²⁰	1	3	3	3	2	4	3	1

Items which score relatively high within a given semantic subset may be seen as prototypical clusters. As Taylor puts it: ‘If we know that an entity is feathered, has wings, and can fly, we can state with some confidence that it also lays eggs.’²¹ To this we might add: and that its name, in Hungarian, will not begin with a nasal.²²

To a less striking degree, but in parallel fashion, most Hungarian bird names contain a s(h)ibilant, liquid, or both; but there are exceptions, e.g., *hattyú* ‘swan’, *héja* ‘goshawk’ and the owls (*bagoly*, *kuvik*, *uhu*).²³

Again, compare Taylor: ‘... such perfect correlations are rare. There are cups with no handles (Chinese cups), birds which don’t fly (penguins), cats without tails (Manx cats), chairs which aren’t for sitting on (dentist’s chairs), and so on.’²⁴

Before I embark on the scrutiny of bird and other names a few words concerning meaning, sense, and form and my approach to them will be helpful. I must also say a few words about the terms *iconicity*, *indexicality*, and *sound symbolism*. I shall then treat the concepts of *paronymic attraction* and *lexical architectonics*, which are complexes built up from the primary concepts of iconicity, indexicality, and symbolism. Finally, in a short excursus, I shall examine Hungarian body-part vocabulary in terms of these complexes.

First of all, *meaning* is to be understood here as a verbal noun with imperfective aspect, that is, as referring to a process — specifically: a mental operation — in which the name and the sense of a word are associated.²⁵ By *sense* I mean ‘the network of interlocking relations with

¹⁹ Wherein refers to a back vowel (orthographically: *u ú o ó a â*).

²⁰ The low ‘score’ of *szem*, which is of Uralic vintage, suggests that not everything about its formal history is known (its Finnish cognate, *silmä*, is also architectonically aberrant: no other monomorphemic Finnish noun has the vowel profile *i—ä* with a medial cluster). The other apparent outlier, *kar*, is a pre-conquest loan from a Turkic language; see *EWU*, p. 692b.

²¹ Taylor, *Categorization*, p. 50.

²² Of the ninety-odd Hungarian bird names examined in this essay, none has an initial nasal (orthographically: *m n ny*; among avian vocabulary the only exception is the life-form taxon term *madár* ‘bird’ itself, which I discuss separately below). The probability of such a distribution’s occurring, if the forms of these names were truly arbitrary, is vanishingly small, since initial nasals are not aberrant in the Hungarian lexical canon: roots with initial nasal make up roughly 15 per cent of monomorphemic vocabulary. Contrast mammalian names, where initial nasals are overabundant: of the thirty-one forms in my core corpus, nine (29 per cent) have initial nasals: *macska* ‘cat’, *medve* ‘bear’, *nyúl* ‘hare’, *nyest* ‘beech/stone marten’, *Martes foina*, *nyuszt* ‘pine-marten’, *Martes martes*, *nyérc* (and its paronym *nerc*) ‘mink’, *Mustela*, *menyét* ‘weasel’, *Mustela nivalis*, *mormota* ‘marmot’, *mókus* ‘squirrel’.

²³ *Bagoly* did contain a liquid until fairly recently, and behaves today as if it still did, as its accusative variant *baglyot* attests. See Abondolo, *HIM*, pp. 204–09.

²⁴ Taylor, *Categorization*, p. 48.

²⁵ Compare also Charles Sanders Peirce’s definition of *meaning* as ‘the translation of a sign into another system of signs’, quoted by Roman Jakobson in his ‘A Few Remarks on Peirce’, in Stephen Rudy (ed.), *Roman Jakobson: Selected Writings. VII. Contributions to Comparative Mythology. Studies in Linguistics and Philology, 1972–1982*, Berlin, 1985, pp. 248–53 (251).

other names or lexical expressions'.²⁶ I take it as axiomatic that such mental associations exist, and that they are a precondition of the use of words. Second, the shape of the name of a word is a sign. *Qua* sign this shape has (Peircean) indexical features.²⁷ Third, these indexical features, or as I shall term them here indices, are clues to the meanings in which words can participate.²⁸

One way to understand how such clues might work is to think of these indices as themselves participating in a kind of meaning, one that resides in the relations obtaining among them. Meaning of this kind is not what is usually thought of as linguistic, since it resembles more the patterning perceptible in music or other non-representational art.²⁹

The visual representation of the full extent of patterning at this level of complexity would require the use of colour and animated graphics. Even at the simple level of bi-phonemic words of the shape VC, Menzerath had to compress and cut corners in order to present a perspicuous picture:³⁰

²⁶ Richard A. Coates, 'Properhood', *Language*, 82, 2006, 2, pp. 356–82 (363). Compare also John Lyons: 'Sense is here defined to hold between the words or expressions of a single language independently of the relationship, if any, which holds between those words or expressions and their referents or denotata', *Semantics*, 1, Cambridge, 1977, p. 206.

²⁷ Compare Ubaldo Steconi on the indexical features of signs: 'When an object really affects another, by causality or contiguity, we can say that the one stands for the other as an index. A hole on a street sign is an index of the bullet that bored it. All signs that we actually exchange in communication have a large indexical component; for example, they are all tokens and as such to some extent indexes of their types. The word "quindi" as it appears here is the index of an abstract Italian form. An index always has an iconic part, there must be in fact a quality that channels the contact with the object; but its representative quality does not depend on it.' U. Steconi, 'The Foundation of a General Theory of Translation Built on the Semiotics of C.S. Peirce' (unpublished doctoral thesis, University of London, 2006), p. 52. In his Peircean approach to the analysis of a subdomain of Finnish lexis Raimo Anttila writes: 'Selection from an inventory is a basic semiotic process, and this fact is indeed recognized by linguists. But not fully. We are dealing with signs in relation to themselves, their material makeup. The rules of such makeup imply of course *canonical* forms. The term is frequently used by linguists, it is true, but almost invariably without semiotic overtones. Canonical forms tell us about the domains of signs. First of all, even at the language-universal level, each language chooses its own shapes from the "common pool". The resulting shapes have then immediate pragmatic meaning. But within one and the same language hierarchical selection continues to map various semantic axes.' See R. Anttila, 'Meaning and Structure of Finnish descriptive vocabulary', in Robert T. Harms and Frances Karttunen (eds), *Texas Linguistic Forum 5: Papers from the Transatlantic Finnish Conference*, Austin, TX, 1976, pp. 1–12.

²⁸ See Robert Austerlitz, 'Japanische Säugetiernamen', in Irmela Hijiya-Kirschner and Jürgen Stalpl (eds), *Bruno Lewin zu Ehren: Festschrift aus Anlaß seines 65. Geburtstages*, Volume 1, *Japan: Sprach- und Literaturwissenschaftliche Beiträge*, Bochum, 1989, pp. 1–11.

²⁹ This powerful but obscure kind of patterning (one might almost say: beauty) is perhaps what Adorno was driving at when he wrote of Goethe's 'Wanderers Nachtlied', namely that it is so superb 'weil darin nicht so sehr das Subjekt redet — eher möchte es, wie in jedem authentischen Gebilde, durch dieses hindurch darin verstummen — sondern weil es durch seine Sprache das Unsagbare der Sprache von Natur imitiert. Nichts anderes dürfte die Norm meinen, im Gedicht sollten Form und Inhalt koinzidieren, wofern sie mehr sein soll als die Phrase der Indifferenz' (*Ästhetische Theorie*, Frankfurt/M., 1970, p. 114). Dante is on a similar path in *De vulgari eloquentia* 1.16 when he argues that 'the *volgare illustre* is an ideal language that will have to be found not with the ears, but with the mind' and metaphorizes it as a panther whose scent is everywhere but which is nowhere apparent (*pantheram [. . .] redolentem ubique et necubi apparentem*). See Joseph, *Limiting the Arbitrary*, pp. 144–45.

³⁰ *Architektonik*, p. 26. Menzerath went no further; for example he did not take the vital next step of examining in any detail the distribution, within these networks, of semantic or grammatical subsets.

In the articulation of high vowels, the tongue occupies a much larger volume in the mouth than it does for the low vowels. The proprioceptive sensation due to this, reinforced by the amount of contact between the sides of the tongue and the upper molars, is available to all speakers and is probably necessary to achieve a precise articulatory gesture. [...] In this perspective, two different languages may easily use the same phonetic variable (vowel height) to convey the same range of sensations (size), and come up with exactly opposite solutions, both being equally iconic; all they need to do is focus upon different parts of the rich sensation package provided by articulatory gestures, in our case the volume of the tongue instead of the size of the air passage between it and the palate.³⁵

What is more, kinds (2) and (3) can overlap or equal what has been termed *sound-symbolic* or *ideophonic*.³⁶ In this essay I restrict my use of the term *sound-symbolic* to refer to all instances of form-to-function matching, whether iconic, indexical, or both.³⁷

Like its tropic cousin metaphor, iconicity tends to hog the spotlight in discussions in which indexicality (whose cousin is metonymy) pines in the shadow. Yet indexicality is as pervasive in language as in all semiosis.³⁸

In compensation for this neglect, I present here a quasi-random and drastically curtailed list of examples of indexicality, encoded as submorphemic phonological features in lexicogrammar (in many cases we can attribute these to documented or clearly reconstructable historical processes):

(1) Russian morpheme-initial non-high front /e/ and deixis (the *é--* of *éto* ‘this’ stands alone in non-foreign vocabulary), and restrictions on the shapes of grammatical desinences; restriction of consonant alternations to verb, as opposed to nominal, inflection;

English morpheme-initial voiced interdental fricative /ð/ and deixis: *this that there thou* (and *though*) stand alone; contrast non-deictic *thick, thin, thigh, thaw*, with /ð/’s voiceless analogue, /θ/;³⁹

(2) in Mandarin Chinese, onomatopoeic (iconic) vocabulary is invariably associated with the first tone: *guā&guā* ‘croak’, *wāng&wāng* ‘bark’, *yī&yā* ‘squeak’, *hēng* ‘(sound of humming)’, *wēng* ‘(sound of buzzing)’;⁴⁰ while indexical fit of form and function has no such prosodic association, for example: *chóuchú* ‘shilly-shally’, *liàngqiàng* ‘stagger’, *yōngzhōng* ‘too fat to move’.⁴¹ Notice, in passing, that most of Yip’s examples also exhibit alliteration or rhyme.

³⁵ Gérard Diffloth, ‘i: big, a: small,’ in *Sound Symbolism*, pp. 107–14 (hereafter ‘i: big’). While both solutions may be ‘equally’ iconic, they are iconic in different ways. I would also go a step further and say that both kinds of iconicity may be exploited by speakers of the ‘same’ language.

³⁶ The terminology referring to this class of forms is unfortunately superabundant: see ‘i: big’, p. 108, with literature. Hungarian terminology here parallels closely the traditional Japanese: *hangutan-zō* ‘onomatopoeic’ (gloss: ‘sound-mimicking’, compare Japanese *giseigo*, *giongo* ‘imitate sound words’) versus *hangulatfestō* ‘mood-depicting’, compare Japanese *gitaigo* ‘imitate attitude words’; see James A. Matisoff, ‘Tone, Intonation and Sound Symbolism in Lahu: loading the syllable canon’, in *Sound Symbolism*, pp. 115–29 (119).

³⁷ The term *symbolic* is therefore infelicitous, since through the thinking of Peirce, Bühler, and Lacan *symbol* has come to refer to precisely that which is neither iconic nor indexical; see Gérard Genette, *Mimologiques: Voyage en Cratylie*, Paris, 1976, p. 487, note 2.

³⁸ Compare note 28.

³⁹ For these and several other examples, see Roman Jakobson and Linda Waugh, *The Sound Shape of Language*, 3rd edn, Berlin, 2002 (hereafter *Sound Shape*), p. 58.

⁴⁰ Po-Ching Yip, *The Chinese Lexicon: A Comprehensive Survey*, London, 2000 (hereafter Yip, *Lexicon*), pp. 180–82, 203.

⁴¹ See *ibid.*, pp. 184–93. He states explicitly that these examples are not onomatopoeic, but rather ‘have higher goals to attain’ (p. 184). There is a trace of circularity to the distinction: Yip excludes, for example, *hóu*, *háo*, and *páo*, all glossed ‘roar’, as if by definition, from the category of onomatopoeia (and it should be noted these are all written with the ‘mouth’ component).

Mandarin and Hungarian reduplicative compounds index their sound-symbolic (specifically: their iconic or indexical) status. Thus we have Mandarin iconic-phonomimetic *chuā&chuā* ‘tramp, tramp, tramp’ and *dā&dā* ‘clatter of horses’ hoofs’ versus indexical *máng&máng* ‘vast and indistinct’ and *mò&mò* ‘affectionate, loving, amorous’, and Hungarian iconic-phonomimetic *dirneg&dörmög* ‘grumble, growl’ and *locs&locs* ‘slush, sludge, squashy mud’ versus indexical *giz&gaz* ‘all kinds of (entangled) weeds’ and *dúl&fúl* ‘fume with rage’.⁴²

(3) *o*-grade occurred in Ancient Greek religious vocabulary denoting artefacts derived with the suffix =*no*;⁴³

Japanese long vowels occur in kinship terms such as (*o*)-*nee*-(*saN*) ‘older sister’, (*o*)-*baa*-(*saN*) ‘grandmother’;⁴⁴

Finnish given names deviate from other nouns in both phonology and morphology;⁴⁵

Apart from a few affective words (*tii&fiy* ‘bat’, *zooosq* ‘break’) South-Sakhalin Nivkh long vowels occurred in only three areas of the lexicon: clause-final scalar particles (*-ii* ‘n’est-ce pas?’), deictics, and classificatory numerals, for example *mix* ‘two (small round objects)’, *miix* ‘two (days)’;⁴⁶

in proto-Algonquian, reduplication was frequent in bird-names, but not in fish or mammal names.⁴⁷

These three kinds of example are all different, and it is useful to try to distinguish them, at least in theory. We have indexing of (1) a grammatical or reference category (parts of speech; deixis, itself indexical: thus English word-initial /*δ*/ indexes an index), (2) a way of signifying (*modus significandi*), namely *sound symbolism*, and (3) semantic domains (religious terminology, numerals, kinship terms, bird names); all being indexed by the setting of a phonological variable. The categories can coexist, as the Nivkh data amply show.

Unfortunately the indexical mode is often overlooked, or worse, is lumped together with the iconic. Margaret Magnus, whose doctoral thesis is a particularly detailed and far-reaching survey of phonosemantics⁴⁸, often uses the word *iconic* metaphorically, to mean indexical or sound-symbolic:

Individual phonemes and phonetic features are meaning-bearing. They each have a unique semantics which can be identified by first measuring the semantic disproportions within phonologically defined classes of words and then the converse — measuring the phonological disproportions within semantic classes. One finds in this way that every word which contains a given phoneme bears an element of meaning which is absent in words not containing this phoneme.

⁴² See Edwin G. Pulleyblank, *Outline of Classical Chinese Grammar*, Vancouver, 1995, p. 9; and Hugh M. Stimson, *Fifty-Five T'ang Poems: A text in the reading and understanding of T'ang poetry*, New Haven, CT, 1976, pp. 7–8 and *passim*. For Hungarian the treatment by John Lotz (*Das ungarische Sprachsystem*, Stockholm, 1939, pp. 170–84; reissued with preface by Thomas A. Sebeok and textual apparatus by Gyula Décsy, Bloomington, IN, 1988) stands alone as treating reduplicative compounding together, and on an equal footing, with other kinds of compounding.

⁴³ See Pierre Chantraine, *La Formation des noms en grec ancien*, Paris, 1933, p. 198.

⁴⁴ Samuel Martin, *The Japanese Language through Time*, New Haven, CT, 1987, p. 71.

⁴⁵ ‘Finnish’ (see note 8).

⁴⁶ Source: field notes of Robert Austerlitz.

⁴⁷ See William Cowan, ‘Reduplicated Bird Names in Algonquian’, *International Journal of American Linguistics*, 38, October 1972, 4, pp. 229–30.

⁴⁸ Margaret Magnus, ‘What’s in a Word? Studies in Phonosemantics’, NTNU (doctoral thesis, Norwegian Technical and Natural Sciences University, Trondheim, 2001), (hereafter ‘Phonosemantics’), <<http://www.trismegistos.com/Dissertation/dissertation.pdf>> [accessed January 7 2007].

One finds further that the effect of the phoneme-meaning varies with the position that the phoneme bears within the syllable. In addition, one finds that all phonemes which have a common phonetic feature also have a common element of meaning.⁴⁹

At first flush, this sounds like Derek Cooke's ideas about semiosis in music,⁵⁰ but Magnus has experiments to validate her inductions. She points perspicaciously to the importance of position within the word (and morpheme).⁵¹

She seems to come close to singling out the indexical a few paragraphs later:

Since no word can function without all these components, it follows that all word meanings are in part arbitrary and in part predictable from their form. Specifically, the referent determines what the word is. The sound does not directly affect what a word denotes, but what it *connotes*, not what it is, but what it is *like*. That is, just by hearing the sound 'brump' in a language, one cannot predict whether the word refers to a sound or an animal or a verb of motion. But if 'brump' refers to a verb of motion, it will involve an initial breaching of some kind of impediment and a sudden, forceful conclusion.⁵²

Magnus is here referring to the articulatory movements entailed in the pronunciation of a word like <brump>:⁵³ stoppage released through an energetic transitional glide (/r/) into a nasal-plus-stop coda with abrupt, voiceless ending. But she then extends these articulatory-kinetic features, metaphorically, onto other kinds of motion, that is, ones not specifically of the vocal tract and not necessarily evocative of sounds or movements causing sounds ('an animal').

We see correlations of form with function on all linguistic levels. At the phonetic level are purer forms of onomatopoeia, which can border on the extra-linguistic: in this regard Richard Rhodes's discussion of what he terms 'wild vocabulary' is helpful.⁵⁴ At the phonological level we should distinguish at least (a) the iconic/synchronic (acoustic and auditory salience and deployment of certain sound types for key morphological signalling) from (b) the indexical/diachronic (consequences of history). Finally, at the lexicogrammatical level we should distinguish (c) the iconic/synchronic, which is greater at the grammatical end, from (d) the indexical/diachronic, which is chiefly lexical and therefore manifold, nebulous, and vast.

The divergence that occurs when languages grow apart is the meat and potatoes of the historical-comparative method: the kind of Latin we see in Portuguese today is no longer so very much like the kind of Latin we see in Romanian, and at greater time-depth: Hindi is not much like Danish. In parallel fashion but within a language, sound change and sense development occurring over the course of its history cause its words to deviate from their older forms and meanings, bringing about another kind of divergence: words whose names and senses were once closer come to be more remote (think here of doublets such as English *road* : *raid*, *shed* : *shade*, *far* : *fare*, and of synchronically unobvious derivatives like *shovel* [compare

⁴⁹ *Ibid.*, p. 7.

⁵⁰ Derek Cooke, *Meaning in Music*, Oxford, 1959.

⁵¹ And for noting that salience of iconicity, not of indexicality, is inversely proportional to the concreteness of the referent: where she writes: 'Nouns display the [iconic] effect much more weakly than verbs or adjectives and Concrete Nouns display the effect least of all'; 'Phonosemantics', p. 31.

⁵² *Ibid.*, p. 9.

⁵³ Magnus's use of pseudo-orthography here must mean that it is the consonants and not the vowels that are of interest.

⁵⁴ Richard Rhodes, 'Aural Images', in *Sound Symbolism*, pp. 276–92; see particularly p. 279 ff.

shove], *handle* [*hand*], or Swedish *tvål* 'soap' : *tvätta* 'to wash',⁵⁵ or Hungarian *fiú* 'son' : *faj* 'race'). Often the names remain identical (homophony) while the senses diverge considerably: English *gate* : *gait*, *flour* : *flower*; Swedish *fria* 'propose marriage to' : *fria* 'acquit'.

The opposite and complementary process, much less regular and much less studied, is the convergent force of *paronymic attraction*. This is a kind of popular etymology in which words in a language, whether once related or not, come to resemble one another, to a greater or lesser degree, in sense, name, or both; in anglophonic linguistics the idea was introduced by Otto Jespersen.⁵⁶ An egregious example is the French organ-stop term *nasard*, distorted, via *nez*, from *nacaire*.⁵⁷

In the modern literature, thinking along these lines may be traced from Mikołaj Kruszewski (1851–87) and Georg von der Gabelentz (1840–93) through Hugo Schuchardt (1842–1927):

Gabelentz, followed by Schuchardt, detected 'a fruitful concept' in these historically 'false' but synchron[ically] valid etymologies, which are based on mass agreement within a given speech community. Words linked together by both sound and meaning manifest elective affinities (*Wahlverwandschaften*), able to modify the shape and the content of the vocables involved.⁵⁸

We may characterize the *architectonics*, or relevant formal properties of the Hungarian lexicon as follows. Hungarian monomorphemic core denotative vocabulary is overwhelmingly mono- or bisyllabic; pronominal and other deictic roots, and interjections, are predominantly monosyllabic. Under one kind of analysis — with the application of a set of mild assumptions and morphonological rules — all Hungarian verb stems may be seen as ending in a consonant. Nouns (and adjectives) are more heterogeneous in shape, but here, too, there are restrictions. For example, there is only one monosyllabic noun ending in a short vowel: *fa* 'tree, wood'; compare *te* 'second person singular pronoun', *le* 'down', *na* '(so) there!', *ne* 'don't!'.⁵⁹

Nominal inflection types correlate, by and large, with types of stem auslaut, or rather: given a set of assumptions and morphonological rules, they may be seen so to correlate.⁵⁹ But there is a more pervasive correlation between sense and form. Stated baldly, it is this: culturally canonic concepts are expressed by phonolexically canonic word-shapes, and conversely, culturally less canonic concepts are expressed by word-shapes which cleave less to the phonolexical canon.⁶⁰

Within certain gross formal subsections of the lexicon, certain prosodic and segmental features tend to outweigh others in lexical frequency, textual frequency, or both. Thus, for example, among Hungarian nominal roots of the shape CVCVC in which both the medial

⁵⁵ Largely replacing older *tvål* 'to wash'. Note that English *towel*, made ultimately of the same etymological stuff as Swedish *tvål*, is isolated in the lexicon.

⁵⁶ In chapter 20 of his *Language: its Nature, Development, and Origin*, London, 1922. Compare also Menzerath, *Architektonik*, p.127: 'begrifflich zusammengehörige Wörter [...] können sich zwar, der engen Assoziation wegen, lautlich beeinflussen.'

⁵⁷ See Pierre Hardouin, 'Essai d'une Sémantique des Jeux de l'Orgue', *Acta Musicologica*, 34, 1962, ½, pp. 29–64, 47.

⁵⁸ See Jakobson and Waugh, *Sound Shape*, pp. 182–83.

⁵⁹ See Abondolo, *HIM*, pp. 17–20.

⁶⁰ Compare Austerlitz, who examined Japanese mammal names in the light of '[...] die Feststellung, daß sich von der Architektur her, d.h. von der schlichten phono-morphologischen Gestaltung einer Vokabel, Schlüsse über Bedeutung, Etymologie und Kultur ziehen lassen, und umgekehrt: semantischer (kultureller) Inhalt kündigt an, daß die eine oder andere Gestaltung zu erwarten ist' ('Säugetiernamen', p. 1.)

and final consonant are obstruents, there is a tendency for the medial obstruent to be more lenis⁶¹ than the final obstruent in non-foreign vocabulary. Thus we have medial stop and final fricative in foreign *fókusz* ‘focus’, *lótusz* ‘lotus’, *topáz* ‘topaz’ and *tubus* ‘tube’, but the inverse distribution in non-foreign *fazék* ‘pot’, *küszöb* ‘threshold’, *kuvik* ‘the scops owl (*Athene noctua*)’, and *fészek* ‘nest’. This phonolexical pattern is one small facet of the overall architectonics of Hungarian, and since it is of the lexicon, it allows of ample exceptions.⁶² For example: *doboz* ‘box’, *koboz* ‘lute’, *toboz* ‘pine-cone’, and *mókus* ‘squirrel’ are not foreign. On the other hand it could be argued that *koboz* and *toboz* belong to a distinct (alternating) stem (and therefore inflectional) class and thus do not align squarely with words like *fókusz*; and the semantics of *mókus* locate it at the periphery of an affective and taboo lexical network. In parallel fashion, in the change of pattern from **losos* to *lázac* we may descry a move, however slight, closer to Hungarian architectonic canon, with the medial voiced fricative *-z-* more lenis, on two counts, than the final voiceless affricate *-c*.⁶³

Examination of a test corpus⁶⁴ of Hungarian body-part names reveals that formal properties are not randomly distributed.⁶⁵ The shapes CVC and CVCC, though frequent in the Hungarian lexicon as a whole, are rare among body-part names: we have only *kar* ‘arm’, *nyak* ‘neck’, *szem* ‘eye’, *comb* ‘thigh’, *test* ‘body’. Far better represented are stem-types with final low vowels alternating with zero, for example: *fej* ‘head’, accusative *feje-t* (and not **fej-t*).⁶⁶ This kind of stem, though textually frequent, is lexically rare in Hungarian. But it makes up an impressive proportion of the core body-part nomenclature: *fej* ‘head’, *fül* ‘ear’, *haj* ‘human head-hair’, *háj* ‘fat’, *máj* ‘liver’, *mell* ‘breast’, *váll* ‘shoulder’, *toll* ‘feather’, *szarv* ‘horn’, *derék* ‘waist’, *fenék* ‘bottom/buttocks’, *bél* ‘gut’, *hát* ‘back’, *ín* ‘sinew’, *íny* ‘gums’, *kéz* ‘hand’, *láb* ‘foot/leg’, *méh* ‘uterus’, *száj* ‘mouth’, *szív* ‘heart’, *szárny* ‘wing’, *árny* ‘shadow’, *térd* ‘knee’, *nyál* ‘saliva’ (twenty-four of eighty terms = 30 per cent). Thus in the domain of body parts we have

⁶¹ For our purposes we may assume for Hungarian a scale ranging from relatively lenis consonants (liquids and nasals) to relatively fortis ones (stops), with the fricatives between these extremes. The status of the affricates is not entirely clear: at least for the formal subset (CVCVC) we are considering here, they seem to hover between the fricatives and the stops. Criss-crossing this stricture-based hierarchy is one of voice: *ceteris paribus* a voiced segment is defined as more lenis than its voiceless counterpart.

⁶² On the complexity of the lexicon and the corresponding complexity of approach required for its investigation see Michael Maratsos, review of Paul Bloom, *How Children Learn the Meanings of Words, Language*, 81, June 2005, 2, pp. 495–98, especially: ‘vocabulary acquisition was never a good candidate in the first place for being a hedgehog problem. Vocabulary provides the chief point at which the enormous variety of human concepts first receives translation into publicly available form (grammar finishes the job, by allowing combinations of words to encode more concepts out of this initial stock)’, p. 497.

⁶³ The lenis-fortis canon is observed in all Hungarian CVCVC non-foreign nominals with both medial and final fricatives, for example: *tavas* ‘spring’, *ravas* ‘cunning’, *kuvas* ‘Hungarian breed of sheep-dog’, *kovás* ‘leaven’. Both fricatives are voiced or unvoiced only in foreign vocabulary: *hasis* ‘hashish’, *tífusz* ‘typhus/typhoid’, *dizőz* ‘female café/cabaret singer’.

⁶⁴ Only core non-compound vocabulary was included. Thus *kéz+fej* HAND+HEAD ‘back of the hand’ and even *szemölcs* ‘mole’ and *félhám* ‘cuticle’ were excluded, along with obscenities and euphemisms, but *tojás* ‘egg’, *farok* ‘tail’ were included). The corpus of eighty terms included the hypernym *test* ‘body’, and a few terms designating parts of non-human bodies, for example, *szárny* ‘wing’.

⁶⁵ For non-arbitrariness in body-part terms see Russell Ultan, ‘Descriptivity Grading of Finnish Body-part Terms’, *Arbeiten des Kölner Universalien-Projekts*, 16, 1975, and ‘Descriptivity in the Domain of Body-part Terms’, *Arbeiten des Kölner Universalien-Projekts*, 24, 1976.

⁶⁶ Termed hereafter *sesquisyllabic* stems.

lexically common CVC and CVCC scarcely represented, while lexically rare sesquisyllabic stems are extremely common.⁶⁷

This striking skew in synchronic distribution is largely the consequence of history. Nearly all stems of the *fěj* type are old, and nearly all CVC and CVCC stems are relatively young. Two principles of diachronic linguistics operate here conjointly. On the one hand, sound change is by and large regular. So for example if in a given language X word-initial *p*- changes to *f*-, provided the sound change runs its course, working its way through the lexicon *p*- will change to *f* in all words, so that eventually all the words that used to begin with *p*- come to begin with *f*-.⁶⁸

On the other hand, in certain language-contact situations certain semantic (and grammatical) classes of word are less likely to be borrowed than others, and, as a corollary, certain vocabulary items are more likely to be retained with the match of name to sense relatively intact. For example, core kinship terms and names of the principal body parts tend to be retained rather than being replaced by borrowings.

Taking these two principles together, we can deduce that in language X core kinship and body-part terms will tend not to have initial *p*-, since all or most such terms are retentions from an earlier time, predating the change of initial *p*- to *f*-. Since both English and Hungarian are languages in which *p*- has changed to *f*-, we can cite concrete examples, namely English *father*, *foot* and Hungarian *férj* 'husband', *fěj* 'head'.

In the remainder of this essay I survey the core Hungarian ornithonyms, mainly from a synchronic perspective. At the end of the section I sketch an overview of Hungarian bird names as a phonolexical set, contrasting them, as appropriate, with ichthyonyms and designations of mammals and parts of the body.

The order in which the terms are considered is determined partly by form, partly by meaning. The format of first presentation is: number (in square brackets) for cross-referencing internal to this essay, the form (in boldface), English gloss, Linnaean binomen,⁶⁹ and date of first sure attestation as an ornithonym (as opposed to use as personal name or toponym). For ease of reference, subsequent citations of a name bear their number in curved brackets.

[1] **csóka** (jackdaw, *Corvus monedula*), c. 1395. The Hungarian name of this medium-sized corvid is in many ways emblematic of Hungarian ornithonyms in general. Of the roughly⁷⁰ ninety bird names in our core corpus, nine (c. 10 per cent) have initial *cs*-, and seven (c. 7.7 per cent) end in the sequence *-ka*; these figures take on significance in light of the corresponding frequencies in the monomorphemic lexicon overall, where the corresponding figures are c. 3.9 per cent and c. 3.25 per cent.⁷¹ The prosodic make-up of the name as a whole also

⁶⁷ The word shape CVCVC considered above in connexion with overall Hungarian lexical architectonics, in which both C_2 and C_3 are obstruents, is also scarcely attested in this semantic subset.

⁶⁸ For theory, data, and literature on lexical diffusion, see William S.-Y. Wang (ed.), *The Lexicon in Phonological Change*, The Hague, 1977.

⁶⁹ These are from Kiss *Madárnevek*; they were checked against Jürgen Nicolai, Dettlef Singer, and Konrad Wothe, *Birds of Britain & Europe*, London, 1994 (translated and adapted from the German edition [Munich, 1993] by Ian Dawson; hereafter *Birds*), and the exhaustive László Gozmány, *Vocabularium Nominum Animalium Europae Septem Linguis Redactum*, 2 vols, Budapest, 1979.

⁷⁰ 'roughly' because of the open approach adopted here to the notion of corpus and to quantitative method in general, as mentioned above. In some connexions a given bird name may be seen as core, in other connexions it may be seen as peripheral.

⁷¹ These counts are based on a corpus of monomorphemic of 4,415 nouns, based primarily on the data screened from Ferenc Papp (ed.), *A magyar nyelv szóvégműtató szótára* [Reverse-alphabetized dictionary of the Hungarian language], Budapest, 1994.

conforms well with the avian canon: of the ninety core bird names, seven (7.8 per cent) have the shape CVCv;⁷² this percentage is higher than the monomorphemic noun corpus as a whole, with 197 (that is, only 4.4 per cent).

Just as striking, and statistically more cogent, are the features lacking from this name: (1a) it does not end in a geminate consonant, or (1b) begin with a consonant cluster (few bird names do), (2) it is not sesquisyllabic (few bird names are). All these facts reflect the history of the language. Word shapes of kinds (1a) and (1b) are typical of vocabulary more recent than avian designations, while sesquisyllabic words tend overwhelmingly to be quite old, and very few bird names date back to Finno-Ugric (let alone Uralic) times.

Finally, we should note that the Hungarian names for all the Hungarian corvids are back-vocalic bisyllables with a heavy first-syllable prosody (long vowel, or short vowel plus coda). Thus alongside *csóka* we have [2] **szajkó** (jay, *Garrulus glandarius*), c. 1525 (underived *szaja*, c. 1395, is now obsolete), [3] **szarka** (magpie, *Pica pica*), ?1217, c. 1395, [4] **varjú** (rook/crow, *Corvus*), c. 1395, and [5] **holló** (raven, *Corvus corax*), c. 1395.

It is striking that five of the eleven bird names that are back-vocalic bisyllables with heavy first-syllable prosody are names of corvids. All the other ornithonyms with names of this shape, with the exception of (11) *banka*, refer to aquatic birds: [6] **szárcsa** (coot, *Fulica atra*), c. 1395, [7] **bakcsó** (night heron, *Nycticorax nycticorax*), 1820, [8] **kócsag** (egret, *Egretta*), 1528, [9] **cankó** (sandpiper/ruff/redshank/greenshank, *Tringa*), 1898, [10] **hattyú**⁷³ (swan, *Cygnus*), c. 1395, and [11] **banka** (hoopoe, *Upupa epops*), 1702. Within this sub-group⁷⁴ there is considerable self-similarity, chiefly in the form of recurring phonemes and phoneme pairs:

sz	á	r	cs	a	
b	a	n	k	a	
c	a	n	k	ó	
			k	ó	cs a g
b	a	k	cs	ó	
h	a	tt	y	ú	

If the first syllable of (7) *bakcsó* is indeed from *vak* ‘blind’,⁷⁵ we might reckon with paronymic attraction away from (4) *varjú* and toward (21) *bagoly*, another nocturnal bird. The generic (9) *cankó* is a conscious creation, a wilful distortion of the dialect form *cakó* ‘stork’.⁷⁶ Notice that the specific distortion involved here, namely insertion of a nasal in the first-syllable coda, renders the form more typical of a bird name: in my core corpus six other terms are of the shape CvCCV: (7) *bakcsó*, (34) *küllő*, (10) *hattyú*, (5) *holló*, (4) *varjú*, (2) *szajkó*; only one is CvCV: (59) *rigó*. Another prototypical cluster may be seen as radiating from *szajkó*: (2) *szajkó* — (7) *bakcsó* — (3) *szarka* — [12] **szerkő**, etymologically a derived doublet to *csér* ‘tern’,⁷⁷ [13] **harkály** (woodpeckers, *Dryocopus*, *Dendrocopos*), c. 1560, [14] **sirály**⁷⁸ (gulls, *Larus*), 1793. Schematically:

⁷² Whereby lower-case <v> represents a short vowel, and upper-case<V> a long.

⁷³ Etymologically identical with (67) *gödény* ‘pelican’, discussed below.

⁷⁴ Excluded from the core corpus as birds not native to, or occurring in, Hungary were *batla* ‘ibis’, and *lunda* ‘puffin’; also the derived *csuszka* ‘nuthatch’.

⁷⁵ *EWU*, p. 71.

⁷⁶ *TESz*, p. 411.

⁷⁷ There are no dedicated philological data for *szerkő* in *EWU* or *TESz*. See Kiss, *Madárnevek*, p. 179.

⁷⁸ Supposedly derived from *sír* ‘cry, weep’ (*EWU* 1332), but included here because of its paronymic patterning with other, non-derived bird names.

ba kcsó
 szajk ó
 szerk ő
 szark a
 hark ály
 sir ály

A third cluster centres on (8) *kócsag*: [15] **gólya** (stork, *Ciconia*), 1585, [16] **kuvik** (scops owl, *Athene noctua*), c. 1795 (but with initial hushing affricate: <tsúvik>), [17] **guvat** (water rail, *Rallus aquaticus*), c. 1795, giving

gólya
 kócsag
 ku vik
 gu vat

The initial *k-* (from earlier *cs-*) of (16) *kuvik* appears to be due to paronymic attraction: with the sole exception of (1) *csóka*, all bird names with initial *cs-* are front-vocalic.

The Hungarian words for ‘crane’, ‘sparrowhawk’, ‘partridge’ and ‘owl’ exhibit mild self-similarity: [18] **daru** (crane, *Grus grus*), c. 1395, [19] **karvaly** (sparrowhawk, *Accipiter nisus*), c. 1395, [20] **fogoly**⁷⁹ (grey partridge, *Perdix perdix*), c. 1395, [21] **bagoly** (various medium-sized strigiform owls), c. 1395:

da ru
 karvaly
 fo goly
 ba goly

The following terms cluster around the presence of a labial followed by a high (or, in the case of the *é* of *gébics*, non-low) front vowel:⁸⁰ [22] **bíbic** (lapwing, *Vanellus vanellus*), c. 1430, [23] **gébics**⁸¹ (various shrikes, *Lanius*), 1793, [24] **pityer** (various pipits, *Anthus*), 1772,⁸² [25] **pinty** (chaffinch, brambling, *Fringilla*), 1533, [26] **fürj** (quail, *Coturnix coturnix*), c. 1395:

bíbi c (&)
 gébi cs
 pi tyer
 pinty
 fürj

⁷⁹ We may add here *hors série* the name of the mythical (totemic) bird [39] **turul**, first attested c. 1282. This name, of Turkic origin, is unusual in having initial *t-*, but its overall shape is reminiscent of the *daru* group and its final *l* also accords well with other bird names.

⁸⁰ This is the renowned phonomimetic *pi, reconstructed as the basis of Latin reduplicating *pipiō* > French and English *pigeon*; historically, (20) *fogoly* also belongs to this grouping, as its first-syllable *o* was originally a high front vowel, to judge from the unproblematic cognates Vakh Khanty *peŋk*, Estonian *püü*, *EWU* p. 403. Reflexes of phonomimetic *pi in Hungarian avian vocabulary include not only the items listed here but also dialect names for the lark: see Kiss, *Madárnevek*, pp. 231–32.

⁸¹ Assumed to have developed from a non-standard form (*Gáborján* or *Gábos*) of the personal name *Gábor*, *EWU* pp. 439 and 452. The shift from back to front prosody required to explain the vocalism of *gébics* (‘Vokalharmonischer Umschlag’, *EWU*, p. 452) accords well with the small size of the shrikes (35–60 grams: *Birds*, pp. 176 and 178).

⁸² Kiss, *Madárnevek*, p. 232.

In parallel fashion, the next grouping is characterized by the presence of the sequence *cs* plus front vowel: [27] **csíz**⁸³ (siskin, *Carduelis spinus*), 1538, [28] **csér**⁸⁴ (common tern, *Sterna hirundo*), 1860, [29] **csicsörke** (serin, *Serinus serinus*; note the reduplication and suffixation), [30] **pacsirta** (various larks, for example, *Alauda arvensis*), 1525, [31] **csirke** (domestic chicken), 1568, giving:

csíz
csér
csicsörke (&)
pacsirta
csirke

An overview of the seven clusters sketched above will assist the reader in grasping the interlocking patterns of recurring partials. Note that certain names, for example (7) *bakcsó*, seem to belong to two clusters equally:

szár <i>csa</i>	da <i>ru</i>	csíz
b ank a	kar va ly	csér
<i>c ank ó</i>	fo goly	csicsörke
k ó csag	ba goly	pacsirta
b akcsó		csirke
h atty ú	<i>g ólya</i>	
	kócsag	<i>gébi cs</i>
ba kcsó	<i>ku vik</i>	bíbi c
szajk ó	<i>gu vat</i>	pi tyer
szerk ó		pinty
szark a	<i>har is</i>	<i>fürj</i>
hark ály	stigl ic	
sir ály	tengel ic(e)	
	bíbic	

Eight bird names seem to cluster as pairs: [32] **túzok** (great bustard, *Otis tarda*), *c.* 1395 and **reznek** (little bustard, *Tetrax tetrax*), *c.* 1600 (dialect word; not in core corpus). These both name bustards, and both are bisyllabics with medial *-z-* and final *-k*. The name of the larger bird is back-vocalic and that of the smaller, front-vocalic (phonometaphoric iconicity). In the pair [33] **szárcsa** (coot, *Fulica atra*), *c.* 1395, and **vércse** (red-footed falcon, *Falco vespertinus*), *c.* 1395 (not in core corpus) we have two forms which are derived, the latter transparently so (from *vér* 'blood'; = *cse* [~ = *csa*] is a denominal noun-forming suffix).⁸⁵

In the pair [34] **küllő** (green woodpecker, *Picus viridis*), 1533, and **csüllő** (kittiwake, *Rissa tridactyla*), 1793, (a dialect word, not in core corpus) the ornithonym *küllő*⁸⁶ is homophonous with an ichthyonym *kiüllő* used to refer to various gudgeons (*Gobio*). The fish name is

⁸³ The *EWU* (p. 221) explains the final *-z* of *csíz* as dissimilation from the *ž* known from the Slavonic forms from which this word is ostensibly borrowed. However in the paronymic context I am trying to elaborate here it is just as likely that a contributing factor was the complete absence of *ž* in (standard) Hungarian bird (and mammal and fish) names, with the sole exception of the reduplicative (50) *zsezse* 'redpoll'. Contrast the exotic *zsiráf* 'giraffe', and note the reduplication and *ž* of the insect name *zsizsik* 'weevil'.

⁸⁴ Detailed discussion of this form, and of its historical connexions with (14) *sirály* and (12) *szerkő* as well as a score of non-standard forms, may be found in *Madármevék*, pp. 178–79.

⁸⁵ *EWU*, p. 1622.

⁸⁶ Earlier and later attestations of this name are associated with different senses: 'bee-eater' (*c.* 1405), 'blackbird' (1838); see *EWU*, p. 857.

relatively recent. It was incorporated into Hungarian piscine nomenclature by Ottó Hermann on the basis of local practice: fishermen referred to these fish as ‘spoke(s)’ (*küllő*) because of the wheelspoke-like appearance of their pectoral fins. Hermann called this manner of naming *alakfestő* SHAPE-PAINTING; in the framework adopted here it is metonymic-metaphoric, since a part of the fish is seen as resembling something else.⁸⁷ Finally, the affricate *c*, the trill *r* and the vowel profile *é—e* of [35] **réce** ‘duck’, *c.* 1395 — this is the generic, but more folksy word; the usual word for the domestic bird is (48) *kacs*, see below — and [36] **jérce** ‘pullet’, *c.* 1533, connect these names with one another and with *vércse*, cited above.

There is a largish cluster of *f*-initial bird names: [37] **fű** (purple gallinule, *Porphyrio porphyrio*),⁸⁸ *c.* 1395, [38] **fajd** (ptarmigans, grouse, capercaillie, *Lagopus*, *Tetrao*, *Tetrastes*), *c.* 1395, (20) *fogoly* ‘partridge’, (26) *fűrj* ‘quail’, [40] **fecske** (swallows, *Hirundo*, *Delichon*), *c.* 1395, [41] **fülemüle** (nightingale, *Luscinia megarhynchos*), *c.* 1395, and [42] **fácán** (pheasant, *Phasianus colchicus*), *c.* 1395. Of these, only *fácán* and *fülemüle* are loans (from Italian and Latin). The others are relatively old — as their initial *f*- (indexically) indicates — save perhaps *fecske*, which is of obscure origin.⁸⁹

The three owl terms show iconicity for relative size:⁹⁰ [43] **uhu** (eagle owl, *Bubo bubo*), 1822, (21) **bagoly** (various owls, *Tyto*, *Asio*, *Nyctea*), *c.* 1395, and (16) **kuvik** (scops owl, *Otus scops*), *c.* 1795. The vowels in the name of the largest owl, *uhu*, are kinomorphomimetic of this bird’s relatively large size; the second-syllable *i* of the name of the smallest owl, *kuvik*, is phonometaphoric, and *bagoly* takes up intermediate position. The final *ly* of *bagoly* links it to (19) *karvaly*, (13) *harkály*, (14) *sirály*, (20) *fogoly*, and (39) *turul*.⁹¹

We may think of bird names such as [44] **haris** (corncrake, *Crex crex*), *c.* 1395 and (11) *banka* ‘hoopoe’ as pivotal in the sense that each has phoneme sequences that reappear as *disiecta membra* in other bird names. The *har-* of (44) *haris* occurs also in (13) *harkály*, and its *-is* resembles the finals of (80) *stiglic*, (79) *tengelic(e)*, (22) *bíbic*, and (23) *gébics*.

The *-ank-* of (11) *banka* occurs also in [45] **danka** (usually *danka+sirály*, black-headed gull, *Larus ridibundus*), 1901, and in (9) *cankó*, and its initial *ba-*⁹² occurs in (21) *bagoly*, (7) *bakcsó*, and *batla*. Though a learned creation of the naturalist Ottó Hermann, *danka* derives from a dialect form found in the works of Bessenyei.⁹³ Finally, five bird names cluster around the vowel profiles *a—a* and *e—e*: [46] **galamb** (various doves, *Columba*), *c.* 1165, [47] **kakas** ‘cockerel’ (generic barnyard term), *c.* 1395, [48] **kacs** ‘duck’ (generic barnyard term), 1548, [49] **geze** (warbler, *Hippolais*)⁹⁴, and [50] **zsezse** (redpoll, *Carduelis flammea*), 1894. Note also, again, the frequent occurrence of velars.

⁸⁷ See János Rác, *A magyar nyelv halnevei*, Budapest, 1996, p. 90.

⁸⁸ This word is labelled as obsolescent by the *EWU* (p. 425). If it is as old as its Khanty cognate (Vakh *pāj*) suggests, it must have designated a different, presumably aquatic, bird.

⁸⁹ Ullmann, when discussing the emotive and aesthetic effects of the collaboration of sound and sense, cites (40) *fecske* as lacking ‘phonetic expressiveness’. His wording is far from clear, but he appears to be thinking along phonometaphoric lines (*Principles*, p. 103).

⁹⁰ Compare also the Finnish designations *huuhkaja* : *pöllö*. The corresponding Swedish pair *uv* : *uggla* does not seem to fit any kind of size-iconic pattern.

⁹¹ No body-part term has final *ly*; among mammal designations we have only *bivaly* ‘buffalo’ and *coboly* ‘sable’.

⁹² Thought to have been influenced by words meaning ‘stupid, silly’ or ‘clumsy’, nearly all of which have initial *ba-* or *ma-* (*mamlasz*, *mafla*, *bamba*, *balek*, *balga*, compare also *ostoba*, *buta*, *mulya*). *EWU*, p. 78.

⁹³ *TESz*, p. 593.

⁹⁴ I was unable to find any philological record for this word.

We may now swiftly survey the remainder of the corpus. As a group these bird names are formally self-similar only to a very slight degree, though the reader will recognize by now several familiar features such as the sequence *ba*, initial *g-*, and a large number of affricates and liquids. The residual corpus includes the following designations of domestic fowl: [51] **csirke** ‘chicken’, 1568, [52] **csibe** ‘chick’, 1757, [53] **liba** ‘goose’, 1565, and [54] **pulyka** ‘turkey’, 1627. In form — bisyllabic non-alternating stems — these names are similar to those of the prime domesticated animals *kutya* ‘dog’ and *macska* ‘cat’. They are not, however, like the alternating stems *ló* (oblique stem *lova-*) ‘horse’, *tehén* (oblique stem *tehene-*) ‘cow’, or *disznó* ([old] oblique stem *diszna-*) ‘pig’, which designate larger and, as it were, more serious domesticated animals. The goose name (53) *liba* resembles both [55] **lile** (various plovers, *Charadrius*), 1786, [56] **lilik** (white-fronted goose, *Anser albifrons*), after 1795, and (61) *lúd*, the more folksy word for ‘goose’. The names [57] **gúnár** ‘gander’, c. 1795, and [58] **gácsér** ‘drake’, 1697, show a submorphemic final *-r* element present also in the male animal names *csődör* ‘stallion’, *ökör* ‘ox’, *kandúr* ‘tomcat’.⁹⁵ The name [59] **rigó** (c. 1395 ‘blackbird’, 1702 ‘golden oriole’)⁹⁶ is without etymology. As we have seen, its shape is not prototypically indexical of a bird name.

Of the residual monosyllabic names some are quite old, namely [60] **sas** (various eagles, *Aquila*, *Hieraetus*), c. 1525, and [61] **lúd** (greylag goose, *Anser anser*), c. 1395, and some are quite new: [62] **sneff** (first attested 1678, a synonym for (71) *szalonka* ‘woodcock, snipe’). The name for the various herons, [63] **gém** (*Ardea*, *Ardeola*), c. 1395, has resisted all etymologizing assaults.⁹⁷ Of Turkic origin are [64] **tyúk** ‘hen’, first attested after 1372, [65] **ölyv** (buzzards, *Buteo*, *Pernis*), c. 1395, [66] **sólyom** (various falcons, *Falco*), c. 1395, and [67] **gödény** (pelican, *Pelecanus*), 1620, the last being a doublet to (10) *hattyú* ‘swan’. Alternating stems among bird names are rare, but we do find [68] **veréb** (various sparrows, *Passer*, *Petronia*), c. 1395, which is from Slavonic (accusative: *verebe-t*), and [69] **vöcsök** (various grebes, *Podiceps*), 1519, which is thought to be Finno-Ugric, with cognates in ObUgrian and Komi (accusative: *vöcsk-öt*).

There are only ten polysyllabic bird names in my core corpus. As a group they stand out as attested relatively late: only ‘vulture’ and ‘turtle dove’ appear before the sixteenth century. We have the mildly self-similar trio [70] **poszáta** (whitethroats, warblers, *Sylvia*), 1590, [71] **szalonka** (snipe, woodcock, *Gallinago*, *Scolopax*), 1742, and [72] **szalakóta** (roller, *Coracias*), 1799;⁹⁸ the vaguely reduplicative trio [73] **kárókatona** (cormorant, *Phalacrocorax carbo*), 1519, [74] **bölgömbika** (bittern, *Botaurus*), 1643, and [75] **gyurgyalag** (bee eater, *Merops apiaster*), c. 1525; and the somewhat isolated [76] **keselyű** (vultures, *Gyps*, *Gypaetus*, *Neophron*), c. 1395. Note also **lappantyú** (nightjar, *Caprimulgus*), 1799⁹⁹ which is not in my core corpus because clearly, if irregularly, derived from *lappan(g)-* ‘be in hiding’.

Three trisyllabic bird names have bisyllabic doublets: [77] **cinege**, a doublet¹⁰⁰ (first attested 1533) to [78] **cinke** (various tits, *Parus*), 1787, [79] **tengelic(e)** (goldfinch, *Carduelis carduelis*), c. 1525, a doublet¹⁰¹ to [80] **stiglic** (first attested 1695), and [81] **gerlice** (turtle dove, collared

⁹⁵ Most Hungarian names for the males of animals end in *-r*, contain the sequence *ka*, or both: note *kan* ‘male animal, especially pig’, (47) *kakas* ‘cockerel’, *bika* ‘bull’ and perhaps even the paronymic *bak* ‘male (especially goat)’ and *kos* ‘ram’.

⁹⁶ Compare German *Amsel* and *Goldamsel*.

⁹⁷ *EWU*, p. 453.

⁹⁸ Kiss, *Madárnevek*, p. 217.

⁹⁹ *Ibid.*, p. 207.

¹⁰⁰ *EWU*, pp. 172–73.

¹⁰¹ *TESz*, p. 887.

dove, *Streptopelia*), after 1372, a doublet to [82] **gerle** (attested from 1830 only).¹⁰² In each case the shorter form is the one attested later. We may see here another example of paronymic attraction, forms which are more indexical of avian sense gradually arising out of forms that are less so and competing and eventually — in the case of *gerle* and *cinke* — prevailing.¹⁰³

The name of the whimbrel, **póling** (not in core corpus) may be seen as pivotal, like (44) *haris* above, in that its initial *pó-* is reminiscent of the *po-* of (70) *poszáta*, and its medial *-li-* of the *li-* of (53) *liba* and (56) *lilik*. The *go-*, *gu-* are perhaps gentle indices of the avian status of [83] **goda** (black-tailed godwit, *Limosa limosa*), 1898,¹⁰⁴ and [84] **gulipán** (avocet, *Recurvirostra avosetta*), for which no philological data were available.

The CVCv prosody of the two bird names [85] **héja** (goshawk, *Accipiter gentilis*), c. 1395, [86] **kánya** (various kites, *Milvus*), 1470, is about all there is indexically avian about them. Similarly [87] **sármány** (bunting, yellowhammer, *Emberiza*), 1685, [88] **seregély** (starling, *Sturnus*), c. 1395, and [89] **sordély** (corn bunting, *Miliaria*), c. 1395, are not prototypical other than through their initial *šVr* sequence. Their relatively low prototypicality is perhaps a reflection of the fact that they are all historically non-monomorphemic, that is, they are motivated vocabulary: *sármány* is historically a compound with the sense YELLOW+BREAST, and both *seregély* and *sordély* are derivatives of themes (or roots built on themes) having to do with bustling motion (*sereg* ‘host, army’, *sürög* ‘bustle’, and the now obsolete *szorog* ‘is worried, busy’).¹⁰⁵ Finally, we have [90] **kakukk** (cuckoo, *Cuculus canorus*), c. 1395, a name of phonomimetic qualities in western Eurasia and elsewhere, but which, with its *ka* and *ku* sequences, aligns rather well with other bird names on the indexical plane as well.

I have mentioned above that the life-form taxon term *madár* ‘bird’ deviates from all Hungarian bird names in having an initial nasal. We may interpret this formal deviance as an index pointing to the fact that historically *madár* is not a bird name at all. This word has no accepted etymology, and to judge from its shape (alternating stem *madár* : *madara-*) it could be quite old. Investigators have sought in vain both in the cognate languages (Uralic) and in potential contact donors (Turkic, Indo-Iranian).¹⁰⁶ If it is old, given its length it must be the reflex of a compound, a derived form, or both. What I propose is that it is of Hungarian origin, that is, it is a motivated, that is, non-arbitrary form. This set of assumptions, in turn, prompts us to reconstruct a form *mVntVr(V), in which it is not difficult to divine a compound consisting of elements known as the present-day Hungarian words *mony* ‘egg’ and *toj-* ‘to lay eggs’, both with cognates throughout Uralic; the syntagm *monyt tojik* ‘lays eggs’ is attested from 1536.¹⁰⁷ We may specify our compound more closely as a taboo circumlocution:

¹⁰² Kiss, *Madárnevek*, p. 193.

¹⁰³ Based on frequency counts given *ss.vv.* in Ferenc Pusztai and Szilvia Csábi (eds), [electronic version of] *Magyar értelmező kéziszótár* [Explanatory desk dictionary of Hungarian], Budapest, 2003.

¹⁰⁴ See Kiss, *Madárnevek*, p. 152, on the exiguous philological record for this word.

¹⁰⁵ See *EWU ss.vv. szorgalom* and *szorog* (p. 450), *sereg* (p. 1320), and *sürög* (p. 1374).

¹⁰⁶ *TESz*, p. 809. The word is first attested as a common noun in the Jókai Codex, which is a mid-fifteenth century partial copy of a manuscript thought to have been written in the last quarter of the fourteenth century. See József Molnár and Györgyi Simon, *Magyar nyelvemlékek*, Budapest, 1980, pp. 67–71.

¹⁰⁷ *EWU*, p. 1525; compare also the cognate Estonian expression *muna too* ‘lays eggs’, cited by the most compendious Uralic etymological dictionary, Károly Rédei (ed), *Uralisches etymologisches Wörterbuch*, Budapest, 1988 (hereafter *UEW*), p. 529. Doubts about the appurtenance of Hungarian *toj-* dissolve when one accepts that the present-day verb is probably an *etymologie croisée*, involving both Uralic and Turkic ancestry. For discussion of other possible compounds in the proto-language see Hartmut Katz, ‘*kainalo*’, in Erhard F. Schiefer (ed.), *Explanationes Et Tractationes Fenno-Ugricas In Honorem Hans Fromm*, Munich, 1979, pp. 109–16.

*mūna+tu(γV)=rV ‘egg-layer’, in which the derivational material reconstructed as *=rV is either agentive or simply a nominalizer found in animal names.¹⁰⁸ Vowel reduction in even-numbered syllables would result in the transitional form *muntVrV which would give, quite regularly, the present-day forms *madara-* ~ *madár*. For the high to low vowel change change *ū > a in derived vocabulary, diagnostic of Hungarian,¹⁰⁹ compare *i > *i > ä (orthographic <e>) in *šijVr(V) > *šijVr(V) > *egere-* ~ *egér* ‘mouse’ and *kūñćV *kūñćV=V > *kuñćV=V > *hangya* ‘ant’ (and probably *kūñćV > *kuñć > *húgy* ‘urine’). Note also Hungarian *facsar-* ‘wring’ from the same *pū(n)ćVr(V)- which gives Finnish *puser=ta-* ‘squeeze, press’ and Hungarian *talál* ‘find, come upon (*invenire*)’ from derived *tūli=IV-, underived *tūli- giving Finnish *tule-* ‘come’.

We now see that the set of Hungarian bird names show, alongside the heterogeneity one would expect of a semantic subset,¹¹⁰ considerable formal homogeneity. Some of this homogeneity is the result of lexical bifurcation (*stiglic* : *tengelic(e)*), and, more obscurely, because of the greater time depth: *hattyú* : *gödény*), some of it is due to differing derivational paths (*csér* : *csirke* : *csicsörke*). And some of it is in all likelihood due, at least in part, to paronymic attraction: the final -z of *csíz*, initial *ku-* (from earlier *csu-*) of (16) *kuvik*, initial *bak-* (from *vak-* ?) of (7) *bakcsó*. The synchronic effect resulting from this history is the self-similarity of bird names, which, taken *en bloc*, functions as an index of what they signify. And it is in this way, in addition to any iconic features they may have, that bird names and presumably most semantic subsets of the lexicon have non-arbitrary features.

The lexicon is vast and shot full of chaos. What I have tried to suggest here is that we will be better able to understand the roles of arbitrariness and its opposites if we think of them as a single and paradoxical complex, having in common ‘with many systems of thought, their status as irresolvable contradictions that catalyze dialectal thought processes and are important sources of creative conflict, resynthesis and reinterpretation’.¹¹¹

Indexicality is studied less than iconicity partly because it arises from causality or contiguity, which latter, oddly enough, is often taken to be the opposite of causality, namely coincidence. But coincidence is not arbitrariness nor is it randomness.

¹⁰⁸ For Samoyedic examples see Toivo Lehtisalo, *Die primären ururalischen Ableitungssuffixe*, Mémoires de la Société finno-ougrienne 72, 1936, pp. 182–83. For Finno-Ugric or Finno-Permic compare for example *šijiri ‘mouse’ (*UEW*, p. 500), *šajVrV ‘nit’ (*UEW*, p. 770), *titiri ‘black grouse’ (with reduplication and suffixation; *UEW*, p. 794).

¹⁰⁹ First presented by Endre Tálos in his ‘Kép szöveg nélkül’, in Gábor Bereczki and Péter Domokos (eds), *Urálisztikai tanulmányok (Hajdú Péter 60. születésnapja tiszteletére)*, pp. 409–20 (418), Budapest, 1983.

¹¹⁰ See Menzerath, *Architektonik*, p. 127.

¹¹¹ Janis Nuckolls, ‘The Case for Sound Symbolism’, *Annual Review of Anthropology*, 28, 1999, pp. 225–52 (246).