

MORPHOLOGICAL DOUBLETS IN POLISH

Peter F. Kipka

Consider the following question. Which would you say: We learned the past tense at primary school or We learnt the past tense at primary school? The question arises because of the existence of doublets such as learned/learnt. This article, which explores the phenomenon of doublets in inflectional morphology, is organised into three sections. The first presents an overview of the framework in which we shall be working, the second introduces a fairly extensive set of data, and the third relates data to theory so that each can cast light on the other.

1. The Framework

In her well known and much replicated study, Berko 1958 clearly demonstrated that when first grade and preschool children use inflectional morphology there is evidence that they have internalised systematic patterns. Thus when presented with a pair of novel creatures each of which has been introduced by the experimenter as a *lun*, *wug* or *tor* (all invented pseudo-nouns), an overwhelming majority of children tested name them *luns*, *wugs* or *tors* respectively. Similar observations were made with respect to verbal morphology such as participle or past tense formation. Given the virtual impossibility of the children concerned simply recalling forms from the ambient linguistic environment, one straightforward interpretation of these results is to say that the children have internalised rules for English word formation (e.g. the pluralisation rule).

In fact it would seem that not only have rules been internalised, but that they are over-generalised, yielding such child utterances as *foots* and *mouses*. Nonetheless, the fact children "recover" from this over-generalisation phase alerts us to the inescapable role played by memorisation during the course of language acquisition.

We are thus apparently confronted by a tension between memorisation on the one hand and the following of patterns or rules on the other. Interestingly, it has become clear that the simplest traditional resolution of this tension does not work. The simplest solution is to claim that one invariably applies the general rule unless one has memorised a specific exception, which may and indeed must pre-empt the rule. Thus we say *luns*, *wugs* and *tors* since nothing has been memorised about such coinages, while **foots* and **mouses* are pre-empted by our having been exposed to *feet* and *mice*. Apart from conceptual problems about the cognitive status of memorised lists, there are empirical objections to this account. Thus in Berko's 1958 experiment, both *heafs* and *heaves* were scored as correct plural forms for *heaf*. In fact, as documented by Pinker & Prince 1988, children at the "over-generalisation stage" often fluctuate between over-generalised and non-over-generalised forms of a lexeme. It is typical to utter both *goed* and *went* over a period of time; the forms co-exist even though it is usually only the former which attracts the attention of parents, observers and diary-keepers. Much the same effect can be demonstrated with adults: there is fluctuation when asked for the past tense of an invented verb like *spling*, but no fluctuation for *splam* (see Bybee & Moder 1983). Furthermore, native speakers of English are quite typically not 100% sure when asked for the past tense of at least a handful of quite frequently occurring verbs. Thus undergraduate students express surprise as they learn that they themselves are uncertain about the "standard" past tense of *learn*, *leap*, *lean* or *lend*. It is striking to find probabilistic response patterns to such verbs from native speakers. Such fluctuation (both at the individual and group level) forces us to conclude that there is no absolute mutual exclusion between memorised morphological knowledge and rule governed morphology.

In fairly recent times, observations such as the above have inspired a family of research paradigms variously referred to as parallel distributed processing, PDP, neural nets or connectionism (see, for instance, Rumelhart, McClelland & the PDP Research Group 1986). This recent line of investigation claims that it is possible to account for both of the two "types" of morphological knowledge mentioned above via a single computer-simulated neural network. The most straightforward such networks have a bank of input units each of which is linked to each of a bank of output units via connections of different strengths (called weights). When any of the input units is switched on, it automatically passes on (to each output unit) this activation scaled or multiplied by the relevant connection strength (typically a fraction). Each output unit totals up all the activation it is receiving at a particular point in time and switches itself on if this total exceeds some threshold. The weights and thresholds can initially be set arbitrarily, with the model then passing through a learning phase prior to final testing. During learning, weights and/or thresholds are incrementally adjusted by a fixed algorithm while a trainer repeatedly presents the model with a set of input-output pairs. The algorithm, which is the technical heart of the model (and of which I suppress the details), might for instance reinforce only those connections which seem to be giving desirable response patterns. In this way, generally after numerous presentations of the training pairs, the model's behaviour approaches its target more and more closely.

It is not difficult to enhance such models with probabilistic features, so that response patterns are given with various degrees of certainty. (Such probabilistic elements could involve, for instance, partial blurring of the input - an assumption that is not totally unrealistic in the context of natural language acquisition.) It can thus happen that two or more responses arise for a single input, sometimes with comparable levels of certainty. This seems desirable and can arise irrespective of whether the training set is a one-to-one or a one-to-many mapping. Crucially, the model automatically generalises beyond the training set, from which it in effect extracts patterns that are applied (probabilistically) to novel input stimuli. It has even been claimed that such models can mirror the intermediate stages undergone by children acquiring their native tongue.

Claims of this nature have triggered much excitement - see Sampson's 1987 evaluation of Rumelhart & McClelland's 1986 model of the acquisition of the English past tense using only the phonetic properties of verbs at the input level. This particular model has not surprisingly been subjected to detailed critiques, the most extensive of which is Pinker & Prince 1988. Pinker & Prince commend the model for its strengths (one of which is its explicit formulation) but also point out a number of undesirable response patterns which they trace to particular design features of the model under consideration. Some of these problems have been addressed in subsequent work (this is in part why I have suppressed a discussion of many of the interesting but intricate implementational details). Others, however, are presented as challenges to the PDP enterprise itself, and not just to one of its implemented instantiations. One such stronger criticism advanced by Pinker & Prince pertains to some of the model's linguistically implausible responses. Some, such as *membled* as the past tense of *mail*, are characterised as "blended hybrids" arising from the model's unselective capitalising on any regularity (linguistically spurious or otherwise) in the training set data. This suggests that linguistic constraints should play a greater role in the design of the model. A further objection pertains to the phenomena of (what are often called) regularity and irregularity. While native speakers of English may hesitate about the past tense of a nonsense verb like *to spling*, they are nonetheless quite certain that it must be *splinged* if they are told that this nonsense verb is regular. While the model is quite (and some might say remarkably) good at capturing probabilistic fluctuation, explaining the lack of such fluctuation where appropriate is not one of its fortes.

Crucially, the Rumelhart & McClelland past tense acquisition model exhibits both strengths and weaknesses. I concur with Pinker & Prince on this point. Significantly, however, the unification of what are traditionally called regular and irregular (or weak and strong) verbs, both being handled by a single homogeneous model, while apparently a strength, could in fact be a weakness - though in an interesting way. Pinker & Prince 1991 entertain the possibility of a mixed model (see also Clark 1989), with traditionally irregular verbs reflecting PDP properties, while traditionally regular verbs are subject to a hard-and-fast categorical past tense rule. Such an approach would naturally embody linguistic constraints which would weed out linguistically implausible innovations; it would also contain both a categorical component and a scalar/probabilistic one. Note that both components are able to generalise beyond a given set of input data. Thus the proposal advanced here is not a simple rules-plus-memorised list model (though the latter can be viewed in mathematical terms as a limiting case of the former). Rather, there are two components and each generalises in its own characteristic way. I am going to explore some of the implications of this idea for the phenomenon of morphological doublets.

Before turning to the data in earnest, I would like to tentatively propose a terminological innovation. The reader may have noticed that I have consistently hedged when referring to regularity and irregularity. I hedge because both types of lexical items (e.g. both weak and strong verbs) enter into generalisations. One could, following Bybee & Slobin 1982, attribute these generalisations to rules and to schemas respectively. For the sake of transparency, however, I propose that we refer to the Unmarked Default Component and the Family Resemblance Component. The former will be invoked to create the past tense of verbs like *talk*, *jump* or *splam*; verbs such as *sing*, *swing*, *sit* or *cut* rely on the overriding power of the latter. One advantage of this approach is that we can now say that all English past tense formation is categorical (there being no fuzziness between the features +PAST and -PAST). Scalarity will enter into the morphological realisation of certain sets of +PAST forms; where it is not sufficiently strong, the default (*ed*-based) pronunciation will emerge.

This approach is grounded in the findings of Bybee & Slobin 1982 and makes fundamental use of the fact that (traditionally called) irregular verbs cluster into families centred on prototypes. The members of these families are all similar along a number of dimensions, with some being similar to most other members, thus being central or near-prototypical, while others are only peripheral members of the family. To give some examples, consider the "no change" class of English verbs (e.g. *cut*, *hit*, *beat*) - these are typically monosyllabic and end in an alveolar stop. Another class, containing *blow*, *throw*, *grow* and *know*, is typified by an initial obstruent-plus-sonorant cluster followed by a diphthong (*fly* is a peripheral member of this class, as evidenced by its past tense form, *flew*). The homophonous verb stems *ring* and *wring* have different past tense forms. This is because two families are implicated here - one contains *sing*, *drink* and *begin* (*ring* belongs here), while the other includes *swing*, *spin*, *stick* (and also *wring*). It appears that this last family is centred on a prototype starting with a consonant cluster and ending in a velar nasal.

It should be clear that family membership conditions are not of the necessary and sufficient sort; rather membership is graded, correlating with degrees of (un)certainly. This contrasts with the linguistic rules of the Default Component, where categorical decisions are permitted to surface. Thus, there are no cases of fluctuating certainty as to whether a "regular" verb ending in [p] should take the past tense ending [t] or [d]. This judgment is hard and fast.

2. Exploring the data

We are now in a position to analyse further examples of doublets using the conceptual framework just discussed. We also need to investigate the adequacy of our conceptual tools with respect to the data. In English, a very substantial set of inflectional doublets contains

alternative plural pairs such as *syllabuses / syllabi* and *curriculumms / curricula*. It is not unreasonable to suppose that use of the more Latinate variants correlates to some extent with knowledge of (some elements of) the classical languages. What is interesting about past tense formation of English verbs is that such straightforward hypotheses (based on factors such as erudition and degree of assimilation of words of foreign origin) are not as easy to generate, suggesting in turn a role for a cognitively based explanation, perhaps along the lines sketched above. Sometimes morphological doublets arise when obsolescent and non-obsolescent forms are in competition (cf. French examples like *assoyant / asseyant* = 'sitting', *je puis / je peux* = 'I can'). Languages richer in inflectional morphology often seem to more readily exhibit doublets. It is thus much easier to find Latin examples (e.g. *explicavi / explicui* = 'I unfolded', *nutrivi / nutrii* = 'I nursed' etc.) merely by opening a standard dictionary than it is to find French ones. In line with this observation, for the remainder of this article I will focus on Polish. Like its Slavic cousins, Polish is inflectionally very rich and so provides us with a wealth of raw material in the form of morphological doublets.

In Table 1, I have counted the number of different doublets found among lexical items beginning with S, T and U in a reasonably comprehensive and up-to-date three volume Polish dictionary (Szymczak 1983). It is, of course, interesting to note the fact that larger Polish dictionaries do list a certain amount of inflectional information for most lexical items. Some of this pertains to declensional or conjugational class membership, but the fact that other information is provided too suggests that for a large proportion of the vocabulary of Polish, mere knowledge of standard inflectional paradigms is insufficient for native speakers. A quick estimate shows that something of the order of 3 to 5% of Polish lexical items exhibit doublets somewhere within their set of inflected forms.

	S	T	U
Genitive <i>j</i>	39	34	5
<i>-i / -ów</i>	52	28	8
<i>-i / Ø</i>	22	8	1
<i>-a / -u</i>	31	12	3
other	16	13	1
Nominative <i>-i / -owie</i>	33	7	1
learned	3	–	1
other	17	4	2
Instrumental	9	3	1
Accusative	10	9	–
Vocative	11	–	2
optionally indeclinable	11	2	1
Adj/Adv	1	1	–
Verbs semelfactive	10	–	12
imperative	4	1	4
other	18	4	6
TOTAL	287	126	48

TABLE 1: **Number of inflectional doublets among entries beginning with S, T, and U in Szymczak 1983.**

Table 1 reveals that nominal declension is a rich source of doublets. Numerous feminine nouns with stems ending in [j] are listed with (obsolescent and very literary sounding) genitive plurals with a final *j*; these can sometimes be heard (more usually they are found in written texts) in place of the usual (final *i*) form:

(1)	Nom.sg.		Genitive plurals
	<i>sekcja</i>	'section'	<i>sekcj / sekcji</i>
	<i>sesja</i>	'session'	<i>sesyj / sesji</i>
	<i>sensacja</i>	'sensation'	<i>sensacyj / sensacji</i>
	<i>spekulacja</i>	'speculation'	<i>spekulacyj / spekulacji</i>

Latinate origin is normal here. Many masculine nouns, however, take either *i* or *ów* genitive plurals (whether they are of foreign origin or not, and irrespective of whether they occur in speech or writing):

(2)	Nom.sg.		Genitive plurals
	<i>sandacz</i>	'pike-perch'	<i>sandaczy / sandaczów</i>
	<i>sandwicz</i>	'sandwich'	<i>sandwiczny / sandwiczów</i>
	<i>skandal</i>	'scandal'	<i>skandali / skandalów</i>
	<i>szal</i>	'scarf'	<i>szali / szalów</i>

Speakers of Polish are not always certain which is (or which are) the "accepted" form for the genitive plural of items like those in (2). In fact, from time to time people will misspeak (often without self-correction), so that one can hear *kluczów* 'keys', where only *kluczy* is "acceptable".

Another set of feminine genitive plural doublets end in either *i* or \emptyset :

(3)	Nom.sg.		Genitive plurals
	<i>szabla</i>	'sword'	<i>szabli / szabel</i>
	<i>stolarnia</i>	'carpenter's shop'	<i>stolarni / stolarni</i>

The last noun in (3) contains a derivational suffix which often produces doublets such as these. I will return to genitive *a / u* doublets under a separate heading, mainly because they enter into numerous sub-patterns that are worth trying to explain; for the moment, a couple examples will suffice:

(4)	Nom.sg.		Genitive plurals
	<i>smakoł yk</i>	'delicacy'	<i>smakoł yka / smakoł yku</i>
	<i>samolocik</i>	'aeroplane (dim.)'	<i>samolocika / samolociku</i>

In the nominative case, masculine personal nouns not infrequently possess two plurals:

(5)	Nom.sg.		Nominative plurals
	<i>senator</i>	'senator'	<i>senatorzy / senatorowie</i>
	<i>sanskrytolog</i>	'Sanskritologist'	<i>sanskrytolodzy / sanskrytologowie</i>
	<i>samuraj</i>	'Samurai'	<i>samuraje / samurajowie</i>

To a certain extent, nominative plurals can give rise to a meaning distinction along affective/social lines. Thus the shorter form tends to come across as more familiar (and it can convey sympathy or jocularity), while the longer form connotes a measure of respect. This is not so clear for the nouns in (5), but is very strongly perceived for those in (6):

(6)	Nom.sg.		Nominative plurals
	<i>chł op</i>	'peasant'	<i>chł opy / chł opowie</i>
	<i>pan</i>	'gentleman'	<i>pany / panowie</i>
	<i>rolnik</i>	'farmer'	<i>rolniki / rolnicy</i>

A few masculine common nouns exhibit both regular (either *i* or *y*) and learned-sounding *a* plurals:

(7)	Nom.sg.		Nominative plurals
	<i>koszt</i>	'cost'	<i>koszty / koszta</i>
	<i>grunt</i>	'land'	<i>grunty / grunta</i>

The pattern in (7) is marginally productive for at least some speakers who can be heard using forms like *numera*, for the plural of *numer* = 'number'. Non-masculine doublets based on other patterns (e.g. *wsi / wsie* = 'villages') also occasionally crop up, both in speech and in dictionaries.

Doublets for some of the seven cases of the Polish declensional system are somewhat more sporadic. We will examine only a few subcases. Masculine accusatives are interesting, as with animate masculines there is accusative-genitive syncretism, while for most non-animate masculine nouns syncretism merges accusative with nominative. The word for pineapple (*ananas*) is worth commenting on. For its accusative singular, it has either *anayasa* (= gen.) or *ananas* (= nom.), perhaps in line with a fruit counting as marginally animate. Furthermore, there is a colloquial use of the lexeme *ananas* to mean something like dunderhead - here animacy is clear and dictates an accusative (*anayasa*) different to the nominative. (Note, however, that it is rare for a fruit to be masculine in Polish. Bananas, which are, behave linguistically like pineapples.)

A handful of nouns exhibit two instrumental plurals, one of which conveys a much more literary flavour:

(8)	Nom.sg.		Instrumental plurals
	<i>oko</i>	'eye'	<i>oczyna / oczami</i>
	<i>ręka</i>	'hand'	<i>ręko / rękami</i>
	<i>ucho</i>	'ear'	<i>uszyna / uszami</i>
	<i>słowo</i>	'word'	<i>słowo / słowami</i>

Even so, a small number of instrumental plural doublets (e.g. those referring to nails or branches) evince no perceptible more literary vs. less literary distinction. Vocative and dative doublets are not frequent, but can be found.

A final point should be made about declensional doublets. While some Polish nouns are indeclinable, thus having one form for all cases (and others, like *muzeum* = 'museum', have a constant form throughout the singular paradigm), there are nouns which are only optionally indeclinable. Some examples are:

- (9) sago
solo
sombbrero
staccato
andante
andantino

Degree of assimilation of lexemes of foreign origin is the obvious factor at work here.

Morphological doublets among adjectives and adverbs are on the rare side, but they do occur (e.g. *tłusty / tłuszczy* = 'fatter', a comparative). Before returning to the promised reconsideration of *a / u* genitives, a few comments about verbs will help to round out the picture. A certain number of semelfactive verbs give rise to doublets. Thus the third person singular past tense of *spuchnąć* = 'to swell' is either *spuchł* or *spuchnął*. Some doublets arise due to the (generally obligatory) phonological rule converting [o] to [u] (spelt *ó*) before

word-final voiced non-nasals becoming (sporadically?) optional. This can happen with imperatives, since the relevant imperative ending is zero - e.g. *sch ł odź / sch ł ódź* for *sch ł odzić* = 'to cool down'. Last of all, it is worth mentioning that doublets arise among those verbs that can belong to more than one conjugational class. Some examples include:

- (10) *kaszleć / kaszlać* 'to cough'
smarkam / smarczę 'I blow my nose'
szemram / szemrzę 'I murmur, grumble'
skamlam / skamlę 'I whine'

It would be interesting to see to what extent a verb's semantics having an acoustic component correlates with the existence of doublets.

3. The *a / u* Phenomenon

However, as foreshadowed above, I will focus on the genitive singular doublets that arise with masculine nouns such as those in (4) - we will call this the *a / u* phenomenon, given that the relevant two genitive singular affixes are *a* and *u*. This phenomenon is of interest because many of the explanations that we have encountered above (e.g. obsolescence competing with innovation; an optional phonological rule; optional indeclinability for a partially integrated noun; semantic connotation; etc.) cannot be successfully invoked in the case of *a / u*. This means that we have a chance to explore the appropriateness of the cognitively based framework developed in the first part of this article. Immediate questions that come to mind include:

- Does the *a / u* phenomenon arise because of the existence of two systems of morphological knowledge (a Family Resemblance Component and an Unmarked Default one)?
- To what extent can we find similarities between properties of past tense formation in English on the one hand and the *a / u* phenomenon in Polish on the other?

These two questions are clearly connected.

Recall that (so-called) irregular past tense formation in English is based on families of verbs. It is worth stating explicitly that (as will be illustrated shortly) many masculine nouns in Polish have a single genitive singular which ends in *a* (call them A-nouns), and another large batch uniquely take *u* (U-nouns). Some nouns allow both possibilities, and native speaker judgments about particular nouns vary - some judgments are robust, while others are at best shaky. An obvious strategy is to try to determine to what extent the relevant nouns fall into families, with central and peripheral members. Within English verb families, phonetic/phonological properties were found to be shared; however, families did not arise via similarities along semantic dimensions. This provides us with an initial hypothesis.

If we consider separately the A-nouns and the U-nouns, it is easy to see that phonologically based families do not seem to be of relevance here. For a start, there are many near minimal pairs, like the following:

- | | | | | |
|------|-------------------------|----------|-------------------------|-----------------|
| (11) | A-noun Genitives | | U-noun Genitives | |
| | <i>s ł upa</i> | 'pole' | <i> ł upu</i> | 'spoils of war' |
| | <i>ogona</i> | 'tail' | <i>zgonu</i> | 'death' |
| | <i>w ł osa</i> | 'hair' | <i>g ł osu</i> | 'voice' |
| | <i>nosa</i> | 'nose' | <i>sosu</i> | 'sauce' |
| | <i>orzecha</i> | 'nut' | <i>oddechu</i> | 'breath' |
| | <i>ko ł nierza</i> | 'collar' | <i> ł upież u</i> | 'dandruff' |
| | <i>ducha</i> | 'spirit' | <i>dachu</i> | 'roof' |

Moreover, sometimes *a* versus *u* in the genitive resolves homophony in the nominative:

(12)	Genitives				
	<i>stracha</i>	'scarecrow'	<i>strachu</i>	'fear'	
	<i>bala</i>	'log'	<i>balu</i>	'ball'	
	<i>zamka</i>	'zip'	<i>zamku</i>	'castle'	
	<i>rocka</i>	'rock (& roll)'	<i>roku</i>	'year'	

This suggests that searching for family resemblance along semantic or other lines is our only viable option. The task will not be trivial given evidence of occasional synonymy like the following:

(13)	Genitives				
	<i>woltomierza</i>	'voltmeter'	<i>woltometru</i>	'voltmeter'	
	<i>wieloką ta</i>	'polygon'	<i>wieloboku</i>	'polygon'	
	<i>skrótowca</i>	'acronym'	<i>akronimu</i>	'acronym'	

It is curious that all the nouns in (13) are examples of quite technical terminology. It is conceivable that the semantic dimensions we are looking for are not easily discriminable in highly technical domains.

Interestingly, however, hundreds of A-nouns do cluster around a number of families based on syntactico-semantic features. Thus personal nouns are all strictly A-nouns, as illustrated in (14), as are animals (see (15)):

(14)	Genitives	
	<i>sanskrytologa</i>	'Sanskritologist'
	<i>senatora</i>	'senator'
	<i>ską pca</i>	'miser'
	<i>sklepikarza</i>	'shopkeeper'
	<i>snoba</i>	'snob'
	<i>studenta</i>	'student'
	<i>stolarza</i>	'carpenter'
	<i>szekspirologa</i>	'Shakespeareologist'

(15)	Genitives	
	<i>skorpionia</i>	'scorpion'
	<i>skunksa</i>	'skunk'
	<i>s\ onia</i>	'elephant'
	<i>s\ owika</i>	'nightingale'
	<i>smoka</i>	'dragon'
	<i>spaniela</i>	'spaniel'
	<i>strusia</i>	'ostrich'
	<i>szczura</i>	'rat'

(16)	Genitives	
	<i>m\ ota</i>	'hammer'
	<i>noż a</i>	'knife'
	<i>widelca</i>	'fork'
	<i>akceleratora</i>	'accelerator'
	<i>zwalniacza</i>	'brake'
	<i>wykrywacza</i>	'detector'
	<i>wy\ą cznika</i>	'switch'
	<i>telewizora</i>	'television set'

Instruments and implements also tend to be A-nouns (see (16)). Perhaps what is at stake here is central (for animates) versus peripheral (for implements) agentivity. This is somewhat reminiscent of the English *er* morpheme which is suffixed onto verbs - sometimes it yields

agents (e.g. *baker, singer*), sometimes implements (e.g. *cooker*), and often it is ambiguous (e.g. *printer, catcher, driver*).

Other families of A-nouns include months of the year (both their present-day names and their archaic or regional ones), meteorologists' terms for cloud types (see (17)), and denominations of currency (both modern and ancient - see (18)).

(17) **Genitives**

altostratusa
altocumulusa
cirrocumulusa
cirrostratusa
cirrusa
cumulonimbusa / u
cumulusa / u
nimbostratusa
stratocumulusa

(18) **Genitives**

dolara	'dollar'
centa	'cent'
denara	'denarius'
talara	'thaler'
funta	'pound'
szterlinga	'pound sterling'
szelą ga	'shilling'
trojaka	[18th century coin]
tynfa	[18th century coin]
<i>grosza</i>	[present day coin]

Next we turn to U-nouns. All four masculine days of the week (i.e. Monday, Tuesday, Thursday, Friday), the names of the senses (sight, hearing, touch, smell, taste), and the words for various noises and sounds (e.g. genitives like *szeptu* = 'whisper', *szczę ku* = 'bark', *syku* = 'hiss', *szmeru* = 'murmur') are examples of U-noun subfamilies. The names of chemical substances (both native and borrowed - see (19)), and most words felt by native speakers to be of foreign origin (see (20)) also constitute families of U-nouns:

(19) **Genitives**

<i>selenu</i>	'selenium'
<i>selenitu</i>	'selenite'
<i>siarczanu</i>	'sulphate'
<i>siarczku</i>	'sulphide'
<i>siarczynu</i>	'sulphite'
<i>skandu</i>	'scandium'
<i>steroidu</i>	'steroid'
<i>strą tu</i>	'strontium'

(20)	Genitives	
	sanskrytu	'Sanskrit'
	sarkazmu	'sarcasm'
	sarkofagu	'sarcophagus'
	sceptycyzmu	'scepticism'
	schematu	'schema'
	segmentu	'segment'
	sejfu	'safe'
	sekansu	'secant'
	sekretu	'secret'
	semestru	'semester'
	senatu	'senate'
	sensu	'sense'
	sentymentu	'sentiment'
	septetu	'septet'
	serwu	'service'
	sloganu	'slogan'
	strajku	'strike'
	<i>sufiksu</i>	'suffix'

One might hypothesise that the suffix *u* is the unmarked default for the masculine genitive singular, as one would expect newly coined and unassimilated words to be more likely to exhibit default behaviour. Genitives like *sfinksa* = 'sphinx' could be viewed as a specific factor (animacy-agentivity, see (15)) taking precedence over the default realisation. Likewise, foreign personal nouns are A-nouns, as are some instruments of foreign origin (e.g. *startera* = 'starter gun', though this line of evidence is not numerically robust). On the other hand it is possible that U-nouns actually do subdivide into family-resemblance groupings. Thus there are doublets for some words of foreign origin (e.g. *sinusu* / *sinusa* = 'sine'), and there is even a family of doublets for nouns perceived to be of Germanic origin beginning with *sz* + C:

(21)	Genitives	
	<i>szkotu</i> / <i>a</i>	'sheet (nautical)'
	<i>szlabanu</i> / <i>a</i>	'barrier'
	<i>szlemu</i> / <i>a</i>	'slam (cards)'
	<i>szpatu</i> / <i>a</i>	'bone spavin'
	<i>sztambuchu</i> / <i>a</i>	'album'
	<i>sztonu</i> / <i>a</i>	'token'
	<i>szwindlu</i> / <i>a</i>	'swindle'

Powers of ten exhibit a pattern that is not really in accordance with a clear default. The information in (22) appears in Szymczak 1983, but one should note that there is speaker uncertainty on this score, and that some dictionaries allow more possibilities of doublets than others:

(22)	Genitives
	<i>milionu</i>
	<i>bilionu</i>
	<i>trylionu</i>
	<i>kwadrylionu</i>
	<i>kwintylionu</i> / <i>u</i>
	<i>septylionu</i> / <i>u</i>

Finally, we turn to diminutives.

(23)	Genitives				
	<i>sklepu</i>	'shop'	<i>sklepiku</i>	'shop (dim.)'	
	<i>soku</i>	'juice'	<i>soczku</i>	'juice (dim.)'	
	<i>spaceru</i>	'walk'	<i>spacerku</i>	'walk (dim.)'	
	<i>zapasu</i>	'supply'	<i>zapasiku</i>	'supply (dim.)'	
(24)	Genitives				
	<i>słupa</i>	'pole'	<i>słupka</i>	'pole (dim.)'	
	<i>swetra</i>	'sweater'	<i>sweterka</i>	'sweater (dim.)'	
	<i>talerza</i>	'plate'	<i>talerzyka</i>	'plate (dim.)'	
	<i>trójkąta</i>	'triangle'	<i>trójkącika</i>	'triangle (dim.)'	
(25)	Genitives				
	<i>stół</i>	'table'	<i>stolika</i>	'table (dim.)'	
	<i>strzępu</i>	'shred'	<i>strzępka</i>	'shred (dim.)'	
	<i>taboretu</i>	'stool'	<i>taborecika</i>	'stool (dim.)'	
	<i>wazonu</i>	'vase'	<i>wazonika</i>	'vase (dim.)'	
(26)	Genitives				
	<i>samolotu</i>	'aeroplane'	<i>samolocika</i>	'aeroplane (dim.)'	
			<i>samolociku</i>	'aeroplane (dim.)'	
	<i>składu</i>	'storehouse'	<i>składzika</i>	'storehouse (dim.)'	
			<i>składziku</i>	'storehouse (dim.)'	
	<i>stosu</i>	'heap'	<i>stosika</i>	'heap (dim.)'	
			<i>stosiku</i>	'heap (dim.)'	
	<i>strychu</i>	'attic'	<i>stryszka</i>	'attic (dim.)'	
			<i>stryszku</i>	'attic (dim.)'	

The data above show that A-nouns seem to always have A-diminutives, whereas U-nouns lead to U-diminutives or to A-diminutives or indeed to both (i.e. to diminutive doublets). The absence of A-nouns with U-diminutives (modulo semantic change) follows from our earlier finding with respect to *a* emerging from a semantically based Family Resemblance Component. The reasoning goes as follows. The semantics of a diminutive comprises the semantics of both the base noun and of diminutivisation. If the semantics of a base noun is sufficient to render it an A-noun, then diminutives constructed on that base will also take *a*. We are assuming, however, that the semantics of diminutivisation does not favour *u*, for if it did the argument would collapse.

In fact, it seems likely that diminutive suffixes (of which there is actually a family in Polish) may create (loosely connected) families of their own, centred around the semantics of smallness, with genitive singulars ending in *a*. The evidence comes from the fact that there are U-nouns with A-diminutives (see (25)). However, there are also U-nouns with U-diminutives (witness (23)). This confirms our suspicion that the component responsible for *u* is not a simple Default Component, but in fact has its own semantically based family resemblance structure. Thus the semantics of U-nouns matters too, and when diminutivising a U-noun, we set up a competition between the base (favouring *u*) and the suffix (favouring *a*). Sometimes the latter wins and an A-diminutive emerges, and sometimes the former wins, triggering a U-diminutive. But with a competition of this sort, things need not be clear cut, and there may be no clear winner. We thus predict that some U-nouns will have diminutive doublets, a prediction that is confirmed in (26).

The data presented appear to be explicable on the assumption that masculine singular genitives in *a* arise from a Family Resemblance Component. The competitor, *u*, may well be

a default, though it does seem to trigger families of its own. In fact, as far as I can see, nothing in principle rules out the possibility that what produces the suffix *u* is both a Family Resemblance Component and the Unmarked Default Component. In fact, I suggest that a Default Component may be primarily a link to another linguistic level (the phonological level in the case of English past tense formation), or it may itself have its own (competing) family resemblance structure.

We can thus conclude that competition between separable components of the morphological system is indeed a fruitful way of analysing at least certain instances of morphological doublets. Because of this competition, there will be occasions when speakers will not be secure in their judgments. Corroboration can readily be found in the statements of pedagogical grammars such as Jaworski 1973:78 warning readers of *a / u* vacillation and enjoining reliance upon dictionaries in such cases. It seems reasonable to view family resemblance components as the way the human cognitive system implements long term "memory", perhaps not so much as an indelible list than as a system for organising information based on patterns evident therein. Certainly one must be careful of statements about morphological operations being mathematical functions (and so yielding unique outputs - see Aronoff 1994:23). One must likewise be equally careful when putting aside certain data (e.g. *dreamed / dreamt*) as "some kind of second-order exception" (Covington 1995:591). What initially appears to be exceptional in one language may well be present far more robustly in another, and can therefore serve to indicate something of the extent to which the human mind functions via competing modules.

REFERENCES

- Aronoff, M. 1994. *Morphology by itself: stems and inflectional classes*. Cambridge Mass.: M.I.T. Press.
- Berko, J. 1958. The child's learning of English morphology. *Word* 14:150-177.
- Bybee, J. & C.Moder. 1983. Morphological classes as natural categories. *Language* 59:251-270.
- Bybee, J. & D.Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58:265-289.
- Clark, A. 1989. *Microcognition: philosophy, cognitive science, and parallel distributed processing* Cambridge Mass.: M.I.T. Press.
- Covington, M. 1995. Review of Hudson. *Language* 71:589-591.
- Jaworski, M. 1974. *Podrę czna gramatyka ję zyka polskiego*. Warsaw : Wydawnictwo Szkolne i Pedagogiczne.
- Pinker, S. & A.Prince. 1988. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73-193.
- Pinker, S. & A.Prince. 1991. Regular and irregular morphology and the psychological status of rules of grammar. In: L.Sutton, C.Johnson & R.Shields (eds) *Proceedings of the seventeenth annual meeting of the Berkeley Linguistics Society 1991*:230-251.
- Rumelhart, D. & J.McClelland. 1986. On learning the past tenses of English verbs In: J.McClelland, D.Rumelhart & the PDP Research Group *Parallel distributed processing: explorations in the microstructure of cognition. Volume 2: Psychological and biological models*. Cambridge Mass.: M.I.T. Press.
- Rumelhart, D., J.McClelland & the PDP Research Group. 1986. *Parallel distributed processing: explorations in the microstructure of cognition*, 2 vols. Cambridge Mass.: M.I.T. Press.
- Sampson, G. 1987. Review of Parallel Distributed Processing. *Language* 63:871-866.
- Stanisławski, J. 1988. *Wielki sównik polsko-angielski*. Warsaw : Wiedza Powszechna.
- Szymczak, M. 1983. *Sównik ję zyka polskiego*. Warsaw : Państwowe Wydawnictwo Naukowe.