

FRED KARLSSON

Frequency Considerations in Morphology

Zusammenfassung

Der vorliegende Aufsatz diskutiert die Rolle von Token-Frequenzen im Rahmen der morphologischen Theoriebildung. Anhand von Fakten aus dem Finnischen wird gezeigt, daß die verschiedenen Wortklassen und Teil-Wortklassen unterschiedliche prototypische Flexionsformen enthalten, die durch die Ständigkeit ihres Auftretens ausgezeichnet sind. Diese Formen haben für die Aneignung und Handhabung, aber offensichtlich auch für die lexikalische Repräsentation von Flexionsparadigmen eine spezielle organisierende Funktion.

1.

Frequency considerations are not part and parcel of standard morphological theorizing or descriptive practice. Mostly, (type or token) frequencies have served only as a pretheoretical tool aiding the linguist in structuring rule systems, inflectional classes, paradigm types, morphophonological alternations, etc. into "major" and "minor" ones. Recently, however, frequency-based arguments have been gaining ground.

Thus, WURZEL (1984), in his monograph on inflectional morphology and naturalness, bases the fundamental concepts of system-defining structural properties as well as inflectional class stability decisively on type frequencies. The system-defining properties of a morphological system are, in instances where there are mutually conflicting candidates, those that clearly outnumber their rivals in terms of type frequency. E.g., it is a system-defining property of the nominal inflection system of Modern Standard German that base forms rather than stems are inflected. There are only a handful of nouns with stem inflections such as *Konto: Kont-en*, *Tuba: Tub-en*, whereas the normal system-defining pattern is base inflection as in *Auto: Auto-s*.

In the same fashion WURZEL characterizes the notion inflectional class stability. Given that there are several partly diverging (sets of) paradigm structure conditions for how words of a particular type could be inflected, that one is most normal that has the greatest number of members in the extant vocabulary. A case in point would be German nouns of the type *Kino/Fresko*. There are some 20 nouns that exclusively form *n*-plurals (*Fresko: Fresk-en*), some 15 nouns with two possible plural forms (*Konto: Konto-s – Kont-en*), but some 100 nouns with exclusive *s*-plurals (*Kino: Kino-s*). The third group is therefore taken to be the inflectionally stable one (id.: 127).

WURZEL's two concepts are explications of the language-specific notions system-directed naturalness (systembezogene Natürlichkeit). This, in turn, is a complement to MAYERTHALER's (1981) universal, semiotically based theory of morphological markedness. Roughly MAYERTHALER's theory predicts that the normal situation is for a se-

manically or pragmatically more complex category to be morphologically symbolized by a more complex form. Plurals should thus be heavier than singulars, past tenses heavier than present tenses, conjunctive moods heavier than indicative moods, etc. It is interesting to note that MAYERTHALER is much more restrictive than WURZEL in assessing the import of frequency considerations. MAYERTHALER (1981: 136ff.) concedes that frequency is an important determinant of the strength of suppletive forms. But otherwise he regards frequency only as an epiphenomenon of naturalness and therefore as more or less devoid of explanatory value. More natural categories tend to be more frequent than less natural categories. The decisive factor is naturalness rather than frequency.

The views thus diverge even in the framework of autonomous natural morphology. Anyway, frequencies have come to the fore also in more behaviorally oriented morphological theories. In psycholinguistics and psychology of language, there is a long tradition of lexico-morphological frequency studies (cf. BUTTERWORTH 1983 for a survey). In fact, frequency effects belong to the most widely attested and undisputed positive results in psycholinguistics. They have been noted in domains as diverse as lexical decision, object-naming, same-different discrimination, and tachistoscope word recognition. Of course, this is not to say that there would exist consensus on how the test results should be theoretically interpreted, or what their eventual significance is for autonomous morphology. Recently, BYBEE & SLOBIN (1982) have accorded token frequencies a conspicuous position in their theory of morphological schemata. The schemata are active in lexical access and they are decisively based upon frequently occurring prototypes. E.g. words ending in /ŋ/ or /ŋk/ are prototypical representatives of a subgroup of English strong past tense forms.

The purpose of the following somewhat tentative remarks is to explore what role *token frequencies* might play in morphology, in particular in paradigmatically oriented morphological theories. In certain respects, the data discussed are relevant for both autonomous and psycholinguistic approaches.

The examples will mostly be drawn from Finnish. As background information, we note that Finnish nominal paradigms comprise some 150 *core forms* (KARLSSON, to appear), i.e. various exponents of the combinations of the morphosyntactic features case, number, and possessive. Normally these are separately symbolized, i.e. without fusion. Nouns have some 1850 more cliticized forms. Finnish verbs have some 530 finite core forms, 320 infinitive forms (including all combinations of cases and possessives), and some 10.000–15.000 participial forms (including all case/number/possessive forms of five participles – this number is possibly even greater, depending upon how inflection and derivation are separated).

2.

The standard definition of a paradigm is roughly that it is a matrix of word-forms structured by all possible combinations of the relevant morphosyntactic features. For morphologically complex languages it is nontrivial to draw a demarcation line between those morphosyntactic features that partake in defining paradigms (typically case, number, gender, tense, person, mood, etc.), and those that don't (especially clitics of various types). This problem shall not occupy us here, however (cf. CARSTAIRS 1984, KARLSSON, to appear, and NEVIS, to appear).

From the point of view of the language system, normal paradigms are by definition full. All their cells are *morphologically* potential word-forms. Very few paradigmatic lacunae are acknowledged on morphological grounds alone. If some cells are totally blocked, the paradigm is defective. Typical examples would be e.g. pluralia and singularia tanta in many languages, or the English modal auxiliaries. But it is often overlooked that the forms of paradigms may also be governed by semantic restrictions. Thus, city and country names, mass words, and certain subclasses of abstract words lack plural forms. Many absolute adjectives have no comparative and superlative forms. "Unpersonal" and meteorological verbs have a restricted set of person forms, etc.

These basically semantic and/or pragmatic restrictions lead to considerable differences in what forms are actually used in various paradigms. The differences are manifested on the level of token frequencies. What is the eventual theoretical and descriptive import of such data?

First consider the data in table 1 (p. 22) which were derived from a computerized corpus of written Finnish containing 26 issues of *Suomen Kuvalehti*, a popular weekly magazine. The number of wordform tokens in the corpus is 610,000. Table 1 shows the token frequencies of the case/number forms of five nouns and one pronoun belonging to different semantic subclasses. *Martti* and *Helsinki* are proper names, *vesi* is a mass word, *alue* denotes place, and *kesä* season, time of the year. These differences in basic meaning predispose the words to occur in particular cases.

The 120 possessive core forms are virtually nonexistent in the corpus and were therefore omitted from the table. Only six possessive word-form types were attested for *alue*, one for *vesi*, none for the other words (cf. the last four lines of the table).

It is striking how the cases are differently favoured in different paradigms. Proper names (normally) and personal pronouns denote human agents. Assuming FILLMORE's classical hierarchy of semantic case marking, such words should be subject-prominent in terms of syntactic behaviour, and theme-prominent in terms of functional sentence perspective. It is therefore to be expected that the nominative, the subject case par excellence, dominates their paradigms. This is also what we encounter in the data. The incidence of nominatives is not less than 94 % for *Martti*, and 59 % for *hän*. (For the sake of simplicity, we base the discussion here and subsequently on the sg. data only). The second major case of proper names and personal pronouns is the genitive, 6 % and 24 %, respectively, which has many syntactic and semantic functions. The dominant function of the genitive in this context is possession. This is also the meaning expressed by the external local cases (adessive, ablative, allative) that constitute the third major use of these words (3 + 1 + 3 = 7 % for *hän*). The notable incidence of possessive case uses (NB: not possessive suffixes!) is also semantically predictable. Possession is typically predicated of humans and their "organizations". The rest of the morphological cases, in particular the internal local ones, are rare in the paradigms of human proper nouns and personal pronouns.

A comparison with the other example paradigms reveals that the genitive recurs with a share of roughly one fourth. The really typical properties of pronouns and proper names are thus a high rate of nominatives and external local cases.

Helsinki is a city name and has locality as one important meaning component. As it denotes the capital of Finland, it is treated as pragmatically given in most discourses. This predisposes it for subjecthood and thematic position, i.e. nominative case

Table 1. Token frequencies of the central case/number forms of the Finnish nominals *Martti* (man's name), *hän* 'he/she', *Helsinki*, *vesi* 'water', *alue* 'region', *kesä* 'summer'. The percentages have been computed separately for the sg. and pl. forms of each paradigm. The accusative is restricted to pronouns. The last four lines provide the total N of tokens of the central case/number forms in the table, the N of the corresponding word-form types, the N of all tokens of the core forms (including possessive forms not listed in the table due to their rareness), and the N of word-form types including possessives.

	<i>Martti</i>		<i>hän</i>		<i>Helsinki</i>		<i>vesi</i>		<i>alue</i>		<i>kesä</i>	
	N	%	N	%	N	%	N	%	N	%	N	%
NOMINATIVE												
Sg.	129	94	2583	59	443	43	31	23	36	14	15	10
Pl.	-	-	-	-	-	-	6	14	13	16	4	57
GENITIVE												
Sg.	8	6	1034	24	372	36	39	29	59	23	42	26
Pl.	-	-	-	-	-	-	9	21	14	18	-	-
PARTITIVE												
Sg.	-	-	181	4	7	1	37	29	20	9	3	2
Pl.	-	-	-	-	-	-	3	7	9	11	-	-
INESSIVE												
Sg.	-	-	15	0	131	13	5	4	-	-	-	-
Pl.	-	-	-	-	-	-	2	9	-	-	-	-
ELATIVE												
Sg.	-	-	63	1	19	2	5	4	4	2	7	4
Pl.	-	-	-	-	-	-	1	2	3	4	-	-
ILLATIVE												
Sg.	-	-	19	0	43	4	11	8	2	1	3	2
Pl.	-	-	-	-	-	-	4	9	4	5	-	-
ADESSIVE												
Sg.	1	1	141	3	2	0	3	2	90	37	56	35
Pl.	-	-	-	-	-	-	9	20	24	32	-	-
ABLATIVE												
Sg.	-	-	26	1	-	-	-	-	10	4	-	-
Pl.	-	-	-	-	-	-	-	7	4	5	-	-
ALLATIVE												
Sg.	-	-	133	3	3	0	1	1	21	9	-	-
Pl.	-	-	-	-	-	-	5	11	7	9	-	-
ESSIVE												
Sg.	-	-	-	-	-	-	-	-	2	1	35	21
Pl.	-	-	-	-	-	-	-	-	-	-	3	43
TRANSLATIVE												
Sg.	-	-	-	-	-	-	-	-	1	0	2	1
Pl.	-	-	-	-	-	-	-	-	-	-	-	-
COMITATIVE												
Sg.	-	-	-	-	-	-	-	-	-	-	-	-
Pl.	-	-	-	-	-	-	-	-	-	-	-	-
INSTRUCTIVE												
Sg.	-	-	-	-	-	-	-	-	-	-	-	-
Pl.	-	-	-	-	-	-	-	-	-	-	-	-
ACCUSATIVE												
Sg.	-	-	179	4	-	-	-	-	-	-	-	-
Pl.	-	-	-	-	-	-	-	-	-	-	-	-
N OF TOKENS												
	138		4374		1020		171		323		170	
N OF TYPES												
	3		10		8		17		18		10	
N OF ALL TOKENS												
	138		4374		1020		177		334		170	
N OF ALL TYPES												
	3		10		8		18		24		10	

(43 %). *Helsinki* also denotes a human conglomerate. This makes understandable the high frequency of (mainly) possessive genitives (36 %). The third major use of this word is in local expressions. The internal local cases (inessive, elative, illative) have a summed incidence of 19 %. Here, the difference is especially clear in comparison to human proper nouns and pronouns.

A common property of pronouns and all proper names is the low predilection (0 %, 4 %, 1 %) to take partitive case. The reason is clear. Partitives typically occur only when the stem is a mass word. Therefore, apart from syntactically triggered partitives such as those under sentence negation, proper nouns and pronouns qualify for partitive use only in the plural.

Vesi 'water' is a mass word. This is reflected in its paradigm by the prominent 29 % share of sg. partitives, clearly the largest one of the words under consideration. Also the internal local cases, sum 16 %, have a visible position. Note in particular how the adessive and allative are surprisingly frequent in the plural (20 %, 11 %) compared to their rareness in the sg. (2 %, 1 %). This is due to some common near-idioms with meanings like '(sail) on the waters' and '(launch a boat) onto the waters'.

Alue 'region' has a relatively indeterminate local meaning and no obvious connections to the primary subject- or theme-enforcing features treated above. Indeterminate locality is expressed primarily by the external local cases which clearly dominate this paradigm. Not less than half of the sg. occurrences of *alue* are in these cases. The incidence of nominatives is only 14 % which reflects the remoteness of agency, thematicness, etc. The internal local cases are rare (3 %, cf. 19 % for *Helsinki*).

Finally, we come to the noun *kesä* 'summer'. It is partly temporal in meaning. Again, we see how word-form occurrence is affected by semantic factors. The dominant cases for expressing temporal meanings are the essive and the adessive, and they certainly rule this paradigm by their 21 % and 35 % shares. Also note that the sg. nominative has a frequency of only 10 %.

The upshot of our argumentation should be obvious. There is in many nominal paradigms a direct relation between basic meaning and frequency of occurrence of individual forms. The meaning components of a word largely predispose what forms of the word are likely to be used. Syntactic and thematic factors also affect how often certain word-form choices will be made. Individual paradigms will thus have different extramorphologically based *frequency profiles*. By this term we refer to the quantitative distribution of the word-forms of the whole paradigm. The dominant forms are the *focal points* of the paradigm. The opposites of focal points are lacunae. As seen in table 1, most core forms (including possessives) are unattested even in large corpora. Some lacunae are due to strict morphological rules. E.g. personal pronouns have no plurals that belong to the same paradigm. Other lacunae represent morphologically potential but semantically more or less anomalous word-forms. E.g. proper names may have plurals in Finnish but they are infrequent and often pragmatically marked when used. (In many ieur. languages there seem to be stricter rules discarding plurals of proper names).

Are semantically induced lacunae morphological analogues of the semantically anomalous but syntactically (perhaps!) wellformed sentences of the classical type *Colorless green ideas sleep furiously*? Sometimes this comparison seems apt. Many of the central meanings of the cases cannot be used if the meaning of the stem is not proper. E.g. the possessive meanings of the external local cases do not fit naturally with tem-

Table 2. Token frequencies of the top 15 word-forms of the verb *olla* 'to be'.

GRAMMATICALLY DEFINED WORD-FORM	N	%	CUMULATIVE %
pres. indic. 3rd person sg.	17224	47,6	47,6
past tense 3rd person sg.	3829	10,6	58,2
pres. indic. 3rd person pl.	3240	9,0	67,2
neg. stem; 2nd person imp. sg.	2752	7,6	74,8
past participle nom. sg.	1483	4,1	78,9
conditional 3rd person sg.	1468	4,1	83,0
1st infinitive nom.	869	2,4	85,4
past tense 3rd person pl.	557	1,5	86,9
pres. indic. 1st person sg.	554	1,5	88,4
pres. participle gen. sg.	387	1,1	89,5
past participle nom. pl.	318	0,9	90,4
pres. indic. 3rd person sg. quest.	294	0,8	91,2
pres. indic. 1st person pl.	291	0,8	92,0
past tense 1st person pl.	145	0,4	92,4
potential 3rd person sg.	142	0,4	92,8
<hr/>			
N OF TOKENS	33533		
N OF ALL TOKENS OF <i>olla</i>	36172		
N OF ALL WORD-FORM TYPES FOUND	180		

Table 3. Recurrence of certain word-form types among the top 15 forms of 12 verbs.
*) *kiinnostaa* 'be interested in'.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	SUM (+)
OLLA																
'be'	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	15
SAADA																
'get'	+	-	-	-	+	+	+	+	+	+	-	+	+	-	+	10
VOIDA																
'can'	+	-	+	+	+	+	+	+	+	+	-	+	+	-	+	12
TÄYTYÄ																
'must'	+	+	-	-	-	+	-	-	-	+	+	-	+	-	-	6
KIINNO- STAA *)	+	+	-	-	+	+	+	-	+	+	-	+	+	+	+	11
TAPPAA																
'kill'	+	-	+	-	+	+	+	-	+	+	-	+	+	-	+	10
ANTAA																
'give'	+	-	-	-	+	+	-	-	+	+	-	+	+	-	+	8
TULLA																
'come'	+	-	-	+	+	+	+	-	+	+	-	+	+	+	+	11
TEHDÄ																
'do'	+	-	+	+	+	+	+	-	+	+	-	+	+	-	+	11
SANOA																
'say'	+	-	+	-	+	+	+	+	-	-	-	+	+	-	-	8
NÄHDÄ																
'see'	+	-	+	-	+	+	-	+	+	-	-	+	+	-	+	9
RAKASTAA																
'love'	+	-	+	+	+	+	-	-	+	-	-	+	+	+	-	9
SUM	12	3	7	5	11	12	8	5	10	9	2	11	12	4	9	

to table 3.

Explication of the columns

1 = pres. indic. 3rd person sg.	9 = negative stem; 2nd person sg. imperative
2 = pres. indic. 3rd person sg. question	10 = pres. conditional 3rd person sg.
3 = pres. indic. 1st person sg.	11 = pres. potential 3rd person sg.
4 = pres. indic. 1st person pl.	12 = 1st infinitive nom.
5 = pres. indic. 3rd person pl.	13 = past participle nom. sg.
6 = past tense 3rd person sg.	14 = pres. participle gen. sg.
7 = past tense 3rd person pl.	15 = past participle nom. pl.
8 = past tense 1st person sg.	

poral nouns. But most cases are so polyfunctional that they have some uses such as government that are not dependent upon proper semantic fit between stem and ending. For this reason, most cells in a paradigm are genuinely potential word-forms.

So far, we have considered only nouns. Verbs are more laborious to analyze due to their wealth of forms, more than 10,000 in Finnish. We shall first restrict ourselves to just one verb paradigm. Table 2 shows the top 15 word-forms of the verb *olla* 'to be' encountered in the 610,000 word corpus.

The copula was instantiated by 180 word-forms in the corpus. Strictly speaking the question form (12th in rank order) is cliticized and not an independent inflected form. The top 15 forms account for not less than 92,8 % of all occurrences. The top five forms account for almost 79 %, and the pres. indic. 3rd person sg. form alone for almost half. This degree of repetition is even stronger than noted above for nominals.

It seems as if verbal paradigms would have more stable focal points than nominals, i.e. that there is less interparadigmatic variation. Consider table 3 which shows how the top 15 forms of *olla* are represented among the top 15 forms of 11 other verbs. Five forms are especially salient and recur in almost all investigated paradigms: the pres. indic. 3rd person sg. and pl., the past tense 3rd person sg., the 1st infinitive nom., and the past participle nom. sg. Table 4 shows in more detail the status of these forms in the respective paradigms.

Table 4. Rank number and relative share of five highly frequent grammatical words in 12 verbal paradigms. Cf. table 3 for definitions of the forms in the columns. For each form, the rows indicate rank number within the paradigm/relative share of all occurrences in the paradigm.

	1	5	6	12	13
<i>olla</i>	1/47,6 %	3/9,0 %	2/10,6 %	7/2,4 %	5/4,1 %
<i>saada</i>	1/19,2 %	5/5,4 %	3/9,9 %	2/10,9 %	4/9,0 %
<i>voida</i>	1/32,4 %	5/5,2 %	4/7,0 %	7/4,2 %	6/4,5 %
<i>täytyä</i>	1/78,7 %	-	3/5,6 %	-	4/3,9 %
<i>kiinnostaa</i>	1/16,0 %	7/5,0 %	6/6,0 %	3/10,0 %	11/3,0 %
<i>tappaa</i>	1/15,1 %	8/3,8 %	9/3,8 %	2/13,2 %	5/5,7 %
<i>antaa</i>	1/13,6 %	10/3,6 %	3/10,5 %	2/13,0 %	4/8,3 %
<i>tulla</i>	1/24,3 %	6/5,0 %	2/15,7 %	5/5,1 %	3/11,4 %
<i>tehdä</i>	4/7,4 %	10/3,4 %	6/5,9 %	1/16,1 %	3/8,3 %
<i>sanoa</i>	1/33,8 %	8/3,0 %	2/14,7 %	3/11,3 %	6/3,9 %
<i>nähdä</i>	1/14,4 %	10/2,8 %	5/5,7 %	2/11,6 %	3/8,9 %
<i>rakastaa</i>	1/11,6 %	3/9,3 %	11/2,3 %	2/11,5 %	11/2,3 %

The most notable thing about table 4 is the almost absolute dominance of the pres. indic. 3rd person sg. form. It clearly tops all paradigms except *tehdä* 'do' which is especially frequent as an infinitival complement. The present indicative 3rd person sg.

form thus has a stronger, less variable, and therefore more salient, position in the verbal inflection system than does the nominative sg. in the nominal system. A second generalization is that the infinitive (12) and past tense (5) forms also have fairly invariable positions among the top word-forms. Of course, there are individual differences among the paradigms that may have syntactic, semantic, or even pragmatic interpretations. E.g. *täytyä* 'must' is monopersonal and therefore has a morphologically and syntactically defective paradigm. *Rakastaa* 'love' favours 3rd person plurals due to its often reciprocal uses, and on the other hand is not very frequent in the past tense, etc. Such facts are more in the nature of idiosyncrasies.

3.

Finally, we shall discuss the theoretical significance of morphological token frequency data. If morphology is seen as a purely autonomous system, such data normally are of little avail. Structural linguistics is above all concerned with the potential of the language system. This is a perfectly sensible enterprise as proven by centuries of grammatical description. In such descriptions, token frequency data are only marginally invoked, e.g. in explanations of the persistence of common suppletive or otherwise irregular forms, or in separating pervasive rules from isolated idiosyncrasies. As noted above, token frequency data also shed some light on where and why there are semantically based paradigmatic lacunae (also cf. PLANK 1981: 128ff.).

When the current data are interpreted from an autonomous point of view, the overwhelming centrality of the pres. indic. 3rd person sg. form in the use of verb forms is especially pertinent. We noted above MAYERTHALER's (1981) claim that token frequencies mostly are epiphenomenal consequences of pervasive markedness principles. On semantic and pragmatic grounds, in particular the principle of speaker centrality, MAYERTHALER postulates that the 1st person is semantically unmarked (unmarkiert) in regard to other persons, and that the 2nd person is unmarked in regard to the 3rd.

But this leads to a conspicuous conflict between markedness status and token frequency in most persons. The 1st person is least marked but fairly infrequent, whereas the 3rd person is most marked and extremely frequent. MAYERTHALER (1981: 30, 44-6) proposes to solve the dilemma by invoking (i) a superordinate universal pragmatic principle, and (ii) the notion of naturalness conflict which states that superordinate principles may override normal markedness relations. According to the universal principle, the speech act participants are perceptually and otherwise so central that it is communicatively beneficial to have them overtly (merkmalhaft) signalled by endings. This leads to a naturalness conflict in the 3rd person. Since it is semantically marked, it should be overtly marked as well, but this is not readily possible due to the already existing marking of the 1st and 2nd persons. The conflict is solved by inverting the overt marking relations and leaving the 3rd person unmarked, or less marked, than the other persons.

In this respect, MAYERTHALER's otherwise compelling argumentation doesn't seem apt. Besides the discrepancy between semantic markedness and frequency data, it also fails to account for why the 1st person is not left overtly unsignalled as semantically unmarked categories (singular, positive, nominative, present tense, indicative, etc.) normally are. No obvious reasons are at hand to explain why the 1st person should be so much more important than other unmarked categories that its surface marking would be understandable.

How, then, should the unmarked status of the 3rd person be motivated? One could e.g. invoke a principle concerning the choice of discourse topic. Speakers normally speak of objects other than the speech act participants. There are literally countless such objects in the universe, but only two speech act participants. Furthermore, language ontogeny shows that 3rd person forms are important from the beginning of language acquisition. And finally, syntactic markedness considerations immediately show that clauses consisting of an NP followed by an appropriate predicate in the pres. indic. 3rd pers. sg. constitute the basic type in many (all?) languages.

Note that this line of argument makes only heuristic use of token frequencies. Here, we agree with MAYERTHALER (1981: 136-40). But such evidence can still be compelling. The normal situation should be for gross token frequency relations to be deducible from markedness principles, which is achievable by postulating the 3rd person as unmarked in the person system.

This solution also has important consequences for how lexical base forms are selected in autonomous descriptions. We would maintain that a corollary of the unmarkedness of the 3rd person is that the stem used in this form is of fundamental importance in a description of verbal inflection. In particular, this stem is more central than the infinitive (stem) which many linguists, e.g. JAKOBSON (1966/1932/) and WURZEL (1984: 53), postulate as unmarked or basic in their descriptions of verbal inflection. Also note that the infinitive syntactically occurs only in complex clause structures involving verb chains, and it is acquired late in language ontogeny. This topic cannot be further pursued here, however.

So far we have discussed the status of paradigm token frequencies in relation to autonomous descriptions of morphological structure. We have stressed the mainly heuristic import of such data. No strong claims have been advanced concerning the direct incorporation or formalization of token frequencies in autonomous descriptions.

The real significance of token frequencies emerges in performance-oriented theories of the perception, production, and acquisition of word-forms and lexical representations. The focal points of paradigms are quantitative stereotypes in the terminology of LEECH and COATES (1980), or prototypes given ROSCH's well-known framework. In other words, they are good exemplars showing how the paradigm is typically instantiated in its normal use(s). To recapitulate some of the Finn. data:

<i>WORD</i>	<i>PROTOTYPICAL FORMS</i>
proper noun [+human]	nom. sg., gen. sg., ext. loc. cases
personal pronoun	- -
proper noun [-human]	nom. sg., gen. sg., int. loc. cases
[+local]	
mass noun	part. sg., int. loc. cases
<i>alue</i> 'region'	ext. loc. cases
<i>kesä</i> 'summer'	ess. sg., transl. sg.
verb	pres. indic. 3rd person sg.

Some of these characterizations concern classes of words, others are tied to individual lexemes. They spell out how meaning properties are reflected in the use of forms. They are not intended to be parts of individual morphological descriptions. Note, however, that all nominal prototypes are not automatically deducible from naturalness considerations which would predict that the nominative sg. is the most frequent form of all paradigms.

The prototypical forms normally are the first to be ontogenetically acquired. Therefore they are especially salient in determining how the child starts building up his/her morphological system. They are also most likely to have multiple or idiosyncratic meaning, and to contain morphological or morphophonological idiosyncracies. We don't claim that token frequencies would be relevant in terms of exact numbers as in the above tables. Such data are, of course, inaccessible to language-users. But on a *gross level* they are reflected in what forms are in recurrent use and thereby real entities for the speaker.

In his review of current theories of the psycholinguistic nature of lexical representations, BUTTERWORTH (1983) notes that the Full Listing Hypothesis (FLH) has not been conclusively falsified. In its strongest form, FLH would claim that all (potential) word-forms of a language are lexically stored as wholes. Literally interpreted this would mean that Finns have mental lexicons that are (tens of) thousands of times larger than English-speaking people's lexicons. This seems highly improbable. The focal point theory provides a moderated version of FLH, viz. that only the prototypical forms are lexically stored. These are the very forms which speakers have heard, actively used, and which make sense pragmatically and semantically. It is likely that many prototypical forms are lexicalized anyway since they are parts of recurrent sayings or idioms that are stored as wholes for semantic reasons. Cf. what was said above of expressions containing *vesi* 'water' in the external local cases. Such a lexicalization theory would account for many apparent peculiar features of the most frequent forms. In particular, it would account for why these forms are readily at hand, easily processable, and first to be acquired. Under any version of FLH, it would also bring the sizes of the lexicons of different languages (e.g. English and Finnish) down to roughly the same magnitude.

This does not mean that there would be no morphological rules, e.g. schemas, word-form generation rules, or inductive rules invoked when the inflectional class of a new item is to be determined. The question of lexical listing vs. rules need not be posed as either-or.

References

- BUTTERWORTH, B. (1983): *Lexical Representation*, in B. BUTTERWORTH, ed., *Language Production*, Vol. 2, New York, 257-94.
- BYBEE, J. L. & SLOBIN, D. I. (1982): *Rules and Schemas in the Development and Use of the English Past Tense*, *Lg* 58, 265-89.
- CARSTAIRS, A. (1983): *Paradigm Economy*, *JL* 10, 115-25.
- (1984): *Constraints on Allomorphy in Inflection*, Indiana University Linguistics Club.
- JAKOBSON, R. (1966/1932): *Zur Struktur des russischen Verbums*, in: E. P. HAMP & al., eds., *Readings in Linguistics*, Vol. II, Chicago, 22-30.
- KARLSSON, F. (to appear): *Paradigms and Word Forms*, in: Roman LASKOWSKI, ed., *Papers from the Workshop on Natural Morphology*, Poznan 1983, Krakow.
- LEECH, G. & COATES, J. (1980): *Semantic Indeterminacy and the Modals*, in: S. GREENBAUM & al., eds., *Studies in English Linguistics for Randolph Quirk*, London, 79-90.
- MAYERHALER, W. (1981): *Morphologische Natürlichkeit*, Wiesbaden.
- NEVIS, J. (to appear): *The Clitic as a Deviation from the Prototypical Word*, Ms. The Ohio State University, Department of Linguistics.
- PLANK, F. (1981): *Morphologische (Ir-)regularitäten*, Tübingen.
- SCHWARTZ, R. G. & TERRELL, B. Y. (1983): *The Role of Input Frequency in Child Language Acquisition*, in: *J. Child. Lang.*, 10, 57-64.
- WURZEL, W. U. (1984): *Flexionsmorphologie und Natürlichkeit*, in: *Studia Grammatica XXI*, Berlin.
- ZWICKY, A. (to appear): *How to Describe Inflections*, Ms. The Ohio State University.