

"DERIVATIONAL" PARADIGMS IN MORPHONOLOGY.

Vito Pirrelli, Stefano Federici

ILC-CNR, Par.O.La sas - Pisa (Italy)

INTRODUCTION.

Traditionally, paradigms were used to deal with inflection in inflectionally rich languages. Only recently (Calder, 1989; Carstairs-McCarthy, 1988, 1992) paradigms have been the object of a far-reaching investigation covering their formal and computational properties. This investigation has highlighted the significance of a paradigm-based treatment of morphological phenomena and its theoretical implications. In this paper, we show how derivational processes in Morphology can be treated **paradigmatically** by using a morphonological network. The approach is not only theoretical speculation but has been subjected to the practical test of a computer implementation. This implementation leads, in our opinion, to a conceptually and computationally cleaner treatment of Morphology.

1 THE PROBLEM.

There are two basic ways to relate a pair of morphonological representations **derivationally**: 1) to take either of them as basic, and derive the other from it via a process of rule-governed phonological change (Fig.1 below); 2) to assume a third representation (somewhat intermediate between the two) as underlying, and make the other two derive from it either by substitution of some phonological segments, or by filling in an underspecified phonological representation (Fig.2 below).

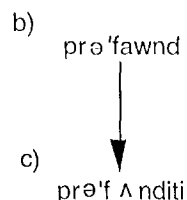


Fig. 1 base-derivative relation in Morphology

Hereafter, we will refer to both 1) and 2) as **derivational** accounts of morphonological relations. Usually, solution 1 stores the basic form only in the lexicon, while the derivative is, as it were, cranked out

on demand by rule application¹. On the other hand, solution 2 requires that only one underlying, abstract representation be stored in the lexicon: the two related forms are yielded from their common lexical source by rules. If b and c are the morphonological representations to relate, the two approaches can be illustrated as shown in Figg. 1 and 2.

In Fig.1, b and c are taken to be on a different footing: the form at the top is more basic than the one at the bottom, the arrow indicating a derivational relation. The following assumption is made: given two forms to relate, it is always possible to specify the direction of the arrow.

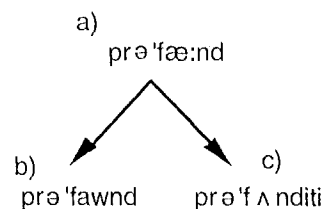


Fig. 2 abstract representations and morphonological relations

This assumption stumbles upon a number of difficulties. Let us consider a small but crucial portion of English segmental Phonology as classically analysed since Chomsky and Halle (1968). First, schwa is well known to alternate with full vowels within the Phonology of English. Derivational series such as in Fig.3 below show that full vowels under stress correspond to schwas in unstressed positions.

[ˈfowtəgræf] / [ˈfowtəˈgræfɪk] / [fəˈtɔ grəfi]

Fig. 3 derivational alternations

This alternation is captured by a rule of vowel reduction: full vowels are reduced to schwa in unstressed position (e.g., /æ/ -> schwa in *photography*). Another systematic phonological alternation in English involves derivational pairs such as *sane-sanity* (['seɪn] / ['sænɪti]), which strongly suggest a derivational rule like eɪ -> æ. However, other alternations appear to go in the opposite direction. For example in a triple like *marginal*

¹Alternatively, both representations can be stored in the Lexicon, and related through the statement of some redundancy lexical rule (Jackendoff, 1975).

/marginlity/ *marginlia*, the following alternation in the underlined vowel emerges: schwa/æ/ej. If we posit /æ/ as underlying, then we end up having to set up an /æ/ -> /ej/ change, which is the mirror-image of the /ej/->/æ/ relation posited for the derivational pair ['sejn]/['sænitɪ] (more on this in Durand, 1990).

The solution illustrated in Fig.2 above is a way to solve this apparent paradox. A third abstract segment /æ:/ is assumed to be basic relative to /schwa/æ/ej/. This means that /schwa/æ/ej/ can be derived from /æ:/ through application of some phonological rules. A derivational chain of this sort can be rather complex, since /æ:/ and /ej/ are, phonologically, far removed from each other, and many intermediate changes can be needed (as in ['sæ:n] -> ['se:n] -> ['se:jn] -> ['se:jn]; see Halle and Mohanan, (1985) for more examples). More reservations on chains of this sort have been expressed recently in a series of psycholinguistic experiments, aimed at probing the reality of the derivational assumptions (Jaeger, 1986; Wang and Derwing, 1986).

In the literature, an altogether different approach from both derivational accounts 1) and 2) above has been suggested (Vennemann, 1974): b and c should simply be listed in the Lexicon, one beside the other, on a par:

prə 'fawnd ←————→ pr ə'f ʌ nditi

Fig. 4 morphological relations and lexical listing

Formulated in this way, the relation in Fig.4 is always "true on the surface", since there is no abstract representation involved in this account. Moreover, no rule ordering problems arise (i.e., concerning the direction of the arrow). If the solution in Fig.4 is adopted, however, it is not at all clear what type of lexical architecture one is suggesting: i.e., it is moot whether b and c are somehow related in the lexicon, or they are simply listed together. More worryingly, the notion "true on the surface" is of little theoretical help for explaining some well-known cases of alternation. Take, for example, the opposition between the American English pronunciation of *writer* and *rider*, respectively /rajɪ'ɾɪ/ and /ra:jɪ'ɾɪ/, where both /t/ and /d/ have been turned into a *flap* ([ɾ]). If the only difference between the two phonological realizations is the surface-true lengthening of /a/ in /ra:jɪ'ɾɪ/, then one is missing the relation between vowel length and the verbal bases of the two derivatives, a relation that represents an important, productive generalization within the American English phonological system (Durand, 1990). An American English speaker, when asked to derive a new agentive in *-er* from verbs ending in /t/, will produce flapping with no lengthening. In contrast, (s)he will produce both a lengthened vowel and a flap if the base is a verb ending in /d/. Derivational accounts capture this generalization in an elegant way, through rule ordering: vowel lengthening takes place before flapping does, in the context "vowel followed by a voiced consonant", so that when either /t/ or /d/ disappears, lengthening has already applied (or failed to). In contrast, accounts based on lexical listing are clearly incapable of grasping this significant hidden relation, since the opposition between /t/ and /d/ is no longer true

on the surface in pairs such as /rajɪ'ɾɪ/ vs /ra:jɪ'ɾɪ/, where the opposition is neutralized by a flap.

To sum up, derivational accounts of morphological alternations face a number of theoretical and practical problems. In what follows, we will illustrate the working of a set of analogy-based principles and the design of a general parallel architecture for their implementation. These principles dispense with both rule ordering and surface untrue representations by using paradigms of alternations instead of derivational chains. The implemented architecture proves to be accurate and computationally efficient.

2 THE GENERAL MODEL.

In this section the idea is illustrated that redundancies among linguistic data can be used to convey interesting linguistic generalizations if data are stored in an incremental network. Let us consider the general case first, exemplified by the following list of abstract cases and their categorial classification:

137	x
037	x
337	x

Suppose that the sequence of numerals 137 represents the form of a linguistic object (say a word); x is its category. We can represent these data by exploiting the redundancies that they show at both the formal and categorial level of description as follows:

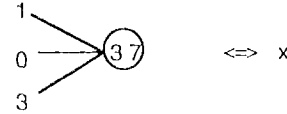


Fig. 5 a formal core and its category

The idea is to interconnect all forms classified the same way by letting them share a common **formal core** if there is any (for a rigorous definition of the notion of formal core, see Pirrelli, 1993). In Fig.5, -37 is the formal core. A core with its category (x in the example at hand) is called a **nucleus**. The graph in Fig.5 can be seen as a network of nodes. Nodes which are linked through a solid line have been witnessed in input as co-occurring in the same form. Nodes which are not linked do not co-occur. Moreover, if two nodes alternate, that is if they are in complementary distribution with respect to a shared core, they occupy the same column in the figure. In other words, nodes on the same column in Fig.5 are mutually exclusive, **paradigmatically related** alternations.

Let us consider now how this network can be used to associate forms with categories. Assume that more data have been stored in the network so that the configuration shown in Fig.6 overleaf is built up. In Fig.6 o is another possible category (different from x), and 38 <=> o another nucleus. Given a network like this, network cores are **activated** by an input string if they are contained by it. The network output will then be the category associated with the activated core.

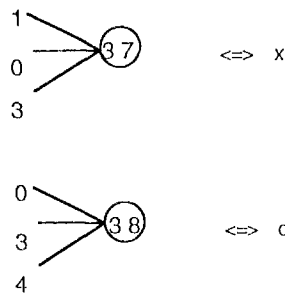


Fig. 6 rival cores in a multicategory network

More concretely, if a certain form, say *437*, is given as input to the network, the system tries to guess the right category on the basis of the analogy of *437* (called the **target string**) with already stored items (called **base strings**). This is carried out through the following steps. Network cores are activated by *437*: this is done through a simple string-matching routine². For example, *37* will be activated, and its category *x* is a candidate response of the system for the input *437*. What happens when more than one core gets activated by the same input token? If the activated cores have the same category, the corresponding multiple responses would reinforce each other: the only category activated is given as output. The case of multiple responses with different categories (and thus potentially different outputs) is more complex. We can distinguish two cases:

- 1) there is no core which is **fully activated**
- 2) there is at least one **fully activated** core.

A core is fully activated when it is entirely contained by the input string: in the example above *437* fully activates the core *-37*. This contrasts with partial activation, when the core is only partially contained by the input string: for example, *437* partially activates the core *-38* of Fig.6 above, since it contains *3* in the second position, but not *8* in the third. Let us consider case 1) above first. If there is no core which is fully activated, then the system goes through the following two steps: **1a)** for each candidate response, the system gauges an **Activation Ratio**:

$$\frac{\text{number of numerals of the activated core contained in input}}{\text{total number of numerals in the activated core}}$$

The activation ratio (AcR) is then $0 < \text{AcR} \leq 1$. The case of full activation (AcR= 1) is dealt with in 2); **1b)** the core with the highest AcR wins out over the others.

Let us consider now case 2). If there exists at least one fully activated core, then two further subcases need be distinguished: **2a)** there is only one core which is fully activated: the category borne out by that core is

² More on string-matching will be said in section 4 of this paper.

then picked up; **2b)** there are at least two cores which are fully activated. 2a) can be seen as a degenerate case of 1) above, when $\text{AcR} = 1$. As to 2b), in this case the activation ratio is clearly no longer conclusive. One needs to gauge a further ratio, called **Analogy Ratio**, whose definition follows:

$$\frac{\text{total number of numerals in the activated core}}{\text{total number of numerals in input}}$$

Again, the analogy ratio (AnR) is $0 < \text{AnR} \leq 1$ ³.

To sum up, a certain pattern is analogically activated if the following conditions are met in this order:

- 1) it is activated by an input pair
- 2) it has got the highest activation ratio
- 3) it has got the highest analogy ratio

So far, we have been assuming that the network produces cores incrementally as more data are input. In fact, the network progressively extracts **reduced** cores. Each full form is already a core in its own right, otherwise it would never get activated. The extraction of smaller cores represents the process of acquiring generalizations on the basis of the analogy between data (redundancies). In the network, an **exception** is simply an isolated case, that is a nucleus which is fully activated only when it is witnessed in input in its entirety (AnR = 1).

3 AN ANALOGY-BASED NETWORK FOR THE SYNTHESIS OF MORPHOLOGICAL ALTERNATIONS.

How is it possible to model the synthesis of the set of systematic alternations illustrated in section 1 by means of a network such as the one in Fig.6? Fig.7-a) overleaf illustrates the result of storing four phonological representations, namely [*vejn*] (*vane*), [*sejn*] (*sane*), [*vænitɪ*] (*vanity*) and [*inæniti*] (*inanity*), plus an (abstract) **categorial feature string** (instead of an atom) for each of them: respectively *xo*, *yo*, *xj* and *zj*. Each categorial feature string is supposed to contain lexico-semantic information (e.g., a capsule representation of the meaning of the relevant lemma) and morphosyntactic information (such as grammatical category, gender, number, tense etc.): the content of the feature string, however, will be relatively neglected here. The machinery of core reduction outlined above obtains for categorial feature strings as well as for their forms. [*ænitɪ*] and [*ejn*] are formal cores; *x*, *i* and *o* are categorial cores.

In Fig.7-b) the arrows between [*s-*] and [*-ænitɪ*], and *y* and *j*, pictures an instance of **paradigm extension**.

³We assume that AnR is calculated only for those cores which have been fully activated (AcR=1). Thus, activated cores cannot contain more numerals than the input does.

Paradigm extension is based on an intuitive idea: if *sane* shares with *vane* one nucleus, then it is expected to share all other paradigmatically-related nuclei.

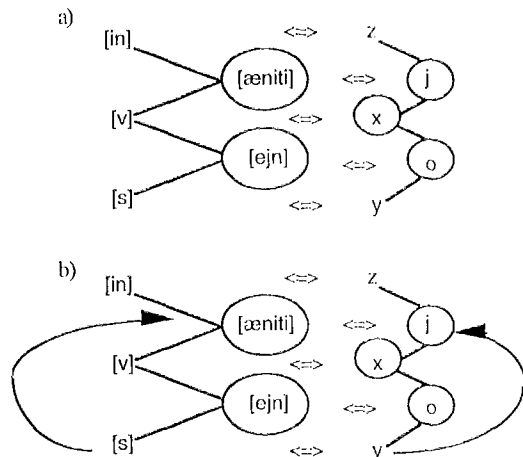


Fig. 7 morphological alternations and paradigm extension

This may not always be true, but is certainly a governing principle in the theory of paradigms. For example, when faced with an unknown *o*-ending adjectival form such as *obsoleto* (English 'obsolete'), an Italian speaker would also predict that *obsoleta* is feminine singular, *obsoleti* masculine plural, *obsolete* feminine plural, according to the paradigm of four-way adjectives in Italian (Matthews, 1992). By the same token, the extension in Fig.7-b) above represents the expectation that the nominalization of *sane* is [ˈsæniiti], by paradigmatic analogy with [ˈvæniiti]. Computationally, this extension is implemented as a continuous path of inter-node connections (say from [s] to [æniiti], through [eɪn] and [v] in Fig.7-b).

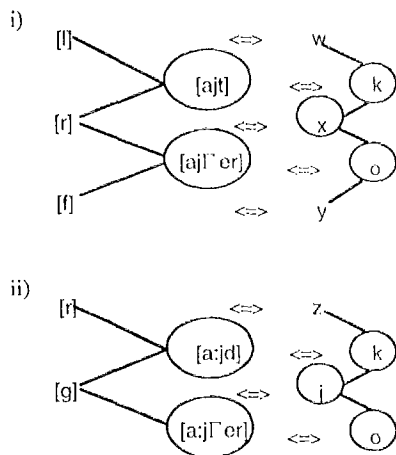


Fig. 8 rival paradigms in a morphological network

Flapping and lengthening, in the American English pronunciation of *writer* and *rider*, can be represented in analogical terms by the two diagrams in Fig.8. There, we have represented the pronunciation of *writer* and *rider* in two separate analogical patterns, respectively i) (without lengthening), and ii) (with lengthening). In fact, i) and ii) should be merged into the same larger pattern. However, it quickly becomes impossible to picture multidimensional links on a page. Fig.8-i is the result of the exposure of the network to the following four word forms: [rajt] (*write*), [rajl'er] (*writer*), [fajl'er] (*fighter*) and [lajt] (*light*). Each word form is given with its categorial feature string: *xk* for [rajt], *xo* for [rajl'er], *yo* for [fajl'er] and *wk* for [lajt]. Fig.8-ii is what the system yields when it comes across the following three word forms: [ga:jl'er] (*guider*), [ga:jd] (*guide*) and [ra:jd] (*ride*). Again, forms are given with their categorial string: *jo* for [ga:jl'er], *jk* for [ga:jd] and *zk* for [ra:jd].

The reader will note that formal cores are parasitic on phonological redundancies, since cores are **extracted** on the basis of systematic surface-true analogies between morphological representations. The difference between suppletion and (semi)regular alternations is easily captured: (semi)regular alternations exploit cores more systematically than do suppletive alternations. Moreover, the structure of the analogical network makes it possible to express hidden phonological constraints as paradigms of phonological alternations. This allows the system to avoid the questionable use of surface untrue, underlying phonological representations, typical of derivational accounts (as in Fig.2 above).

To make the last two points clearer, let us turn to a concrete instance of word synthesis by analogy. A typical objection levelled at representations which treat alternations by listing them, is that crude lists of stored items do not make a clear distinction between regular alternations and irregular ones. Given a raw list of cases, it is objected, nothing can be predicted from such a list, in much the same way as nothing can be predicted from - say - the suppletive alternation *go/vent*.

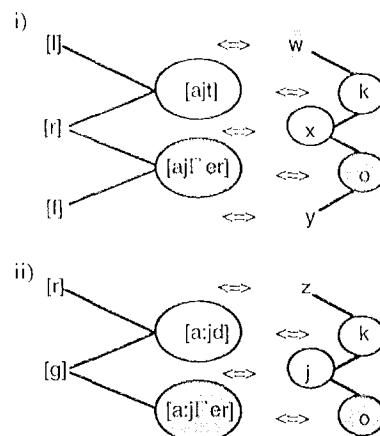


Fig. 9 word synthesis through morphological paradigms

However, paradigmatic extension accounts for the fact that flapping is a productive alternation in American English.

Let us suppose that the system has to produce the agentive nominalization (*lighter*) of ['lajt], whose surface form and lexical content has been already stored in the network (Fig.9-i above). A flap (with no lengthening) would then automatically appear in the place of surface [-t], according to the following steps: a) the input conditions are represented by the lexical content of the verb *light* (*w* in Fig.9-i), and the categorial information "agentive nominalization" (*o* in Fig.9); b) *w* activates the [l] of [lajt] in the paradigm i) of Fig.9 above; c) the activated core *o* triggers the flap alternation [-ajΓer] of word-final [-ajt] in Fig.9-i, and the alternation [ajΓer] of word final [-a:jd] in Fig.9-ii; d) the form ['lajΓer] is thus produced, since there is a path of paradigmatic links between [l] and [ajΓer], while there being no such a path linking [l] and the alternation [a:jd] by paradigm extension.

It should be noted that only (semi)regular alternations share some (sub)core(s) in common. Suppletions such as *go/went* simply do not. Nevertheless, no clear-cut distinction is drawn here between minor alternations and irregular forms, since all of them are simply stored in an analogy-based network. Their difference is accounted for in terms of a gradation, defined by the fact that regular alternations overlap with other forms more extensively than do suppletions. We believe that this solution is empirically superior to the Anaximander's principle invoked by Hudson (1974), to the effect that, since there exists no clear-cut distinction between suppletions and systematic alternations, all of them should be listed in the lexicon.

4 LIMITS OF THE MODEL AND FURTHER IMPROVEMENTS.

The network has been implemented in C, and proved to be 95% accurate in analysis after a training on 20,000 Italian word forms, and 75% accurate in synthesis, as reported in Pirrelli and Federici (1991). These performances have been obtained by using the simplified model illustrated here.

It should be noted, however, that this model works well for the Morphology of those languages (such as Italian and English) where affixation is commonly realized through a concatenative operation (suffixation or prefixation). In these cases, string-matching is a fairly simple head-and-tail operation. Clearly, this model is far from having a universal, cross-linguistic validity though. For example, it does not work well with cases of circumfixation (known also as parasynthesis), let alone the Morphology of nonconcatenative languages. Nevertheless, we contend that the set of analogy-based principles illustrated here hold for a wider spectrum of languages than purely concatenative ones: by making string-matching a more flexible and powerful operation, we can successfully adapt our model to the requirements of nonconcatenative languages such as Arabic, or to the treatment of discontinuous affixation. What follows sketchily illustrates this line of development.

Given two strings of characters/phonemes to match (called the base and the target string), the new string-matching algorithm (**discontinuous string-matching**) we are currently experimenting on is, informally:

start from the left corner of both the target and base string and scan them rightwards; extract, down the way, all characters/phonemes which appear in both, and in the same order.

Take the Italian word forms [inve'k:jare] (*invecchiare*, English 'to age' a parasynthetic verbal derivative of the adjective *vecchio*, 'old'⁴) and [intimi'dire] (*intimidire*, English 'to make shy, intimidate', a parasynthetic derivative of *timido*, 'timid, shy'), or the Arabic pair [kulayb] ('little dog') and [kalb] ('dog') in the diagrams in Fig.10. In the Arabic example, the shared portion of meaning (DOG) triggers the extraction of a tentative core (a stem). By using a discontinuous string-matching, the meaning DOG is associated with the match *k_lb*. The remaining portions of morpholexical features of the two word forms are respectively associated with the substrings left out of the match. Note that the two *a*'s which are linked by a dotted line will **not** be extracted in the same pattern as *k_lb*, since they are ordered differently relative to the position of *l* (in the top string, *a* follows *l*, while in the bottom string, *a* precedes it). In the Italian example, the shared morphological feature *infinitive* triggers the extraction of the circumfix [in__re].

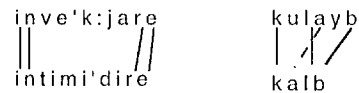


Fig. 10 the use of discontinuous string-matching in Italian and Arabic.

Discontinuous string-matching captures an important range of phenomena which would be treated clumsily by using a simple head-and-tail matching. A further advantage of this routine over even more unconstrained conceivable routines is that it reduces considerably the number of candidate matches. On the more negative side, we are not able to find matches whose order in the base string is interchanged relative to the order in the target, as in English *un-re-do* and *re-un-pack* in Fig.11-a) overleaf. This by no means implies that phenomena such as so-called inversion or metathesis, as in the Rotuman example of Fig.11-b) overleaf, are beyond the reach of our approach. The idea is that a match is extracted only when a base and a target string show the same order of interchangeable characters, as in Fig.11-c) overleaf. The productive morphological relation between, say, *pa* and *ap* in the word 'pigeon' in Rotuman is accounted for **paradigmatically**, as explained in section 3 of this paper. Intuitively, we do not say that a sequence *pa* is transformed into *ap* under

⁴ We depart, here, from Scalise's analysis of *invecchiare* as the product of the sequential application of a suffix [-are] and a prefix [in-] in this order (Scalise, 1983).

particular circumstances. Rather, the sequence *pa* alternates with the sequence *ap* within a certain paradigm.

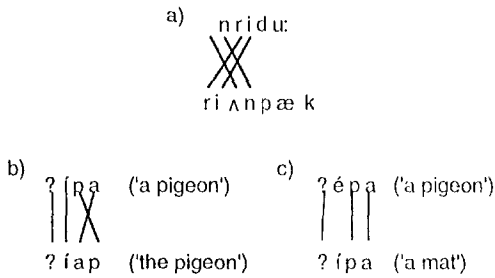


Fig. 11 discontinuous string-matching does not allow for cross-matches.

5 CONCLUSIONS.

The architecture illustrated here can be looked at as a model of the morphological competence of a native speaker, and functions both in analysis and generation. At this point, a question naturally arises: is it a lexicon or a grammar? In our opinion, it is both: a self-modelling lexicon which extracts generalizations on the basis of the analogies between already stored items, and uses these generalizations as repair strategies for lack of direct evidence provided by an already stored item. This interpenetration of lexicon and grammar offers a few advantages in dealing with the problem at hand:

- **phonological concreteness of lexical representations:** known lexical items are given a full surface phonological representation in the lexical network: no abstract phonological segments such as [æ:] in the phonological representation of *profound* in Fig.2 are needed;

- **alternations vs deletions and inversions:** changes in the phonological structure of a given entry are always expressed as alternations, never as deletions; this has the obvious implication that all alternating segments are encoded in the network and can be retrieved at any time: e.g., one does not say that flapping makes a dental disappear, but that flapping can appear only in those paradigmatic contexts where a dental is also present; this limits the computational power of the required rule set considerably, since deletive rules are computationally most costly; moreover, by treating cases of metathesis paradigmatically, the system further spares the computational price of a transformational operation such as inversion.

- **unordered rules:** the context for a given phonological change to take place can be specified in terms of a whole paradigm; this formulation has the immediate advantage of avoiding the need for rule-ordering: for example, the fact that *writer* in the American-English pronunciation has no lengthened [aj] is traditionally accounted for by ordering vowel lengthening before flapping; this move is no longer

needed as long as one can restrict the application of vowel lengthening to the paradigm of those nuclei which present [a:] in the base form of a verb, as we showed above;

- **phonological concreteness of extracted generalizations:** phonological generalizations are expressed in terms of extracted nuclei; nuclei represent what is shared by two (or more) stored items; it follows that no phonological features appear in extracted nuclei other than those which occur in the surface representations from which nuclei are derived;

- **use of direct evidence:** as we saw, smaller nuclei (morphemes) come into play as repair strategies only; that is, if and only if there exists no whole lexical representation (an already known word) which matches either the input (in analysis) or the output (in generation) in its entirety; otherwise, bigger nuclei always override smaller ones; this accounts for phenomena of lexical blocking (Aronoff, 1976), when - say - the word *business* is never interpreted as the 'quality of being busy'.

From a more dynamic perspective the network illustrated above can be seen as an abstract model of the gradual learning of morphological phenomena by a speaker. In fact, the principles illustrated above were originally developed as principles informing machine language learning as such (Federici, 1990), and tested in dealing with some complex cases of unsupervised acquisition of particular linguistic capabilities (e.g., in tasks of bi-lingual translation). Research carried out in this paper proves that morphological phenomena are within the grasp of these learning principles. This represents an important *a-posteriori* confirmation of the validity of the integration between models of machine learning and computational models of linguistic competence, an integration that has already shown its merits in cognitive approaches to language learning and linguistic theorizing (Pinker, 1989; Carstairs-McCarthy and Stemberger, 1988).

REFERENCES.

- Aronoff, M. (1976). *Word Formation in Generative Grammar*, MIT Press.
- Calder, J. (1989). Paradigmatic Morphology. in Proceedings of ACL, Manchester.
- Carstairs-McCarthy, A. (1988). Non-concatenative inflection and paradigm economy. In Hammond M. and Noonan M. Eds, *Theoretical Morphology: Approaches in Modern Linguistics*, pp. 71-77, San Diego: Academic Press.
- Carstairs-McCarthy, A. (1992). *Current Morphology*, London: Routledge.
- Carstairs-McCarthy, A. and J. P. Stemberger (1988). A processing constraint on inflectional homonymy. *Linguistics* 26, pp. 601-17.
- Chomsky, N. and M. Halle (1968). *The Sound Pattern of English*, New York: Harper and Row.
- Durand, J. (1990). *Generative and Non-Linear Phonology*, Longman.

Federici, S. (1990). *Un Sistema Connessionista Auto-Espandibile di Comprensione del Linguaggio Naturale*, Laurea Dissertation, Università di Pisa, Dpt. of Computer Sciences.

Halle, M. and K. P. Mohanan (1985). The segmental phonology of Modern English. *Linguistic Inquiry* 16, pp.57-116.

Hudson, G. (1974). The representation of non-productive alternation. In J. M. Anderson and C. Jones Eds. *Historical Phonology*, 2, Amsterdam: North-Holland.

Jackendoff, R. (1975). Morphological and Semantic Regularities in the Lexicon. *Language*, 51, pp. 639-71.

Jaeger, J. J. (1986). On the acquisition of abstract representations for English vowels. *Phonology Yearbook*, 3, pp. 71-97.

Matthews P. H. (1992). *Morphology* (second edition) CUP.

Pinker, S. (1989). *Learnability and Cognition*, MIT Press.

Pirrelli, V. (1993). *Morphology, Analogy and Machine Translation.*, Phd Dissertation, Salford University UK.

Pirrelli, V. and S. Federici (1991). An analogical way to language modelling: MORPHEUS. *ACTA LINGUISTICA HUNGARICA* 41.

Scalise, S. (1983). *Morfologia Lessicale*, CLESP: Padova.

Vennemann, T. (1974). *Words and syllables in natural generative grammar*. In A. Bruck et al. Eds., *Papers from the parasession on Natural Phonology*, (Chicago,CLS), pp. 346-74.

Wang, H. S. and B. Derwing (1986). More on English vowel shift: the back vowel question. *Phonology Yearbook* 3, pp. 99-116.

Lexicon

