The balance of rules and memory in inflection

An overview by Eric Auer

29. November 2001

In this document, I summarize a paper / book chapter by *Baayen et al.* (*Dutch inflection: The rules that prove the exception*) in which they argue that even clearly regular forms can be handled by memory. However, they assume complex factors determining the balance of rules and memory, rejecting some common explanations for it. Evidence are frequency effects not only for the base lexeme but also for the surface (inflected) word forms. They give some experimental data for the frequency effects with Dutch verbal inflection. First, they test the regularity of Dutch noun plurals using both theoretical considerations and a production study.

1 Introduction: Regularity, defaults and human language

While classical linguistic approaches have been trying to squeeze out every regularity in language (like *Bloomfield* and *Chomsky* did), more diversity about the balance of storage and computation can be found among psycholinguists. *Buttherworth* and connectionists like *Seidenberg* advocate the extreme of most assuming memory as the main device. People like *Pinker*, *Marcus* and *Clahsen* on the other hand love the generalizations that can be drawn from rules, considering lists as unelegant kludges for the really irregular things. Yet others have proposed dual route models, where both rules and memory compete to find a solution in some way, like *Schreuder* and *Baayen*. For implementation purposes, computational linguists are also well aware of possible resource savings by putting *some* common but complex forms into memory, as a shortcut for expensive parsing processes.

As the question of the balance of storage and computation has been around for quite a while, some existing tests and models are discussed by the authors. One of those assumptions is that there are separate components for regular and irregular inflection, and according to Pinker and Clahsen, only irregulars (which are handled by memory) are supposed to show certain frequency effects. The authors have a look at regularity and frequency effects for both Dutch plural nouns and Dutch verbal inflection.

1.1 Two kinds of frequency effects

Frequency effects are generally considered to be distinguishable into *base* and *surface* effects: While Base Frequency Effects are thought as an effect of accessing the underlying *lexeme* (which is the same for all inflectional variants), Surface Frequency Effects can be seen as evidence that the particular inflected *form* itself is represented in some mental storage. So Pinker and Clahsen claim that there should be no Surface Frequency Effects for regular inflection, because their model excludes regular (rule based) inflection from using the memory which is reserved for the irregulars.

Baayen et al. refer to several other researchers who have found such Surface Frequency effects for Dutch, Italian, English and even Finnish. Finnish has a very rich morphology, so one would assume a general human preference to use rules handling this language to avoid overcrowding the memory with too many inflected forms. For the other languages, the bias towards rules seems to be less strong. As *Landauer* has estimated the storage capacity of the brain to be quite impressive, it is not at all clear that the effort put into rule processing is worth the savings on the memory requirements. As rules can also be very expensive in terms of processing time and complexity, the authers put into question the classical approaches to use rules all over the place. They do not, however, ban rules (as some connectionists do). Rules are quite useful when it comes to generalizations or when low frequency but regular processes are to be handled. So the goal is to gain some new insights about the rule/memory tradeoff.

There are also widely accepted cases of storing regular forms: With nous that are used in the plural form most of the time (such as *feet*), it is known that the plural form can be the main instance in the mental lexicon instead of the usual way of deriving the plural from the singular form. There are even languages (like Bari) where a singular suffix exists for derivation of singular forms from such default plural nouns.

1.2 Two kinds of regularity

While the intuition tells us that what is handled by simple rules is called a regular form, Clahsen et al. argue that one has to differentiate mere regular inflection from regular *default* rules. Defending the same point, *Marcus et al.* argue that in German the real (default) regular noun plural is the rarely used *-s* suffix, while the much more common *-en* and other suffixes are taken to be somewhat irregular. Others like Pinker, *Prince* and *Gordon* have on similar reasons questioned the regularity of the Dutch noun plural: The Dutch noun plural is handled by – apart from a few irregular and semi-irregular exceptions – the suffixes *-en* and *-s*.

The authors start the main part of their paper with an analysis whether any one of both has to be considered the default, rendering the other one an irregularity in some way. In that case, irregularity could be taken to be the main criterion that influences the balance of storage to computation. The authors argue that something being handled by a rule does not mean that it (at least frequent instances of it) may not be participating in memorizing processes as well.

As the next section will show, the authors come to the conclusion that in Dutch, *both* plural suffixes are fully regular and productive, so there would be in fact two default suffixes.

2 How regular are Dutch noun plurals?

To give some additional support to earlier experiments where Surface Frequency Effects were shown for Dutch noun plurals, Baayen et al. have to check the regularity of the Dutch noun plural system. For this, they consider the notion of default as used by Marcus et al., and they conduct a production experiment to prove the productivity of both common Dutch plural noun suffixes, *-en* and *-s*. The *-eren* suffix and other exceptional cases are not considered here, they are considered to be handled by listing them.

The selection of the plural suffix in Dutch is based on at least five criteria from various aspects of language: The most prominent is *phonology*, where *-s* is selected after unstressed syllables and *-en* after stressed ones. After a schwa, there is a preference for *-en* but *-s* is also possible. Those are modified by the other factors, an important one being *morphology*, which requires a certain plural form for some but not all common suffixes (e.g. *-tje* requires *s*). There is also a *semantic* influence: Loan words use *-s*, and there is a preference for *-s* with nouns denoting persons (like in *portiers*).

What we have are rival suffixes as described by *Van Marle*: The selection of a suffix is based on several dimensions, and sometimes we end up in a situation where both suffixes are possible if we look at all of those dimensions.

Marcus et al. and Clahsen et al. argue for German that its more common plural suffixes are both not what Marcus and Clahsen call the default, but rather one of the not so common but allegedly more productive one. This is what can be seen as the *elsewhere condition*, as *Kiparsky* uses it and which goes back to *Panini* in some way. It is the one rule that comes up when the other possibilities are too restricted to be applied or fail for some other reason. Using the same argument, one could claim that only *-s* is the default regular noun plural suffix in Dutch, the others being more or less irregular.

2.1 Theoretical considerations about the default

The authors check the Dutch suffixes using the tests used by Marcus et al. for their claim about German. According to the tests, the default is what applies to *new words* (like in a wug-test: *Berko* asked people to give the plural for words he had just invented). For Dutch, *both* suffixes can be used for new words, mostly controlled by the phonological criterion. The default is also used for more *specialized* variations (like in *portiers*) and to *talk about words* (like in: A sentence with two *ofen/schippen* in it). While the former shows some preference for *-s*, the latter is again controlled by phonology (the rhythmic principle), thus allowing *both* suffixes.

The position of the plural suffixes for *non-canonical roots* can be disputed: For example for loan words, *-en* would collide with the suffix commonly used with load verbs, and for words borrowed from English, the *-s* suffix is related to the use of the same suffix in the English plural. Acronyms and surnames, which are also non-canonical roots, can take both suffixes. Lacking some concrete data, the authors give way to their intuition when they assume both suffixes (controlled by phonology) are used in cases of memory failure such as in speech errors.

So at the end, the authors come to

the conclusion that there is no strong evidence for *-s* being the default noun plural suffix in Dutch. To strengthen their point, they test the productivity of both suffixes, as full productivity is said to be only available for default rules. In other cases, as with English *-s/-z/-iz*, nobody has yet tried to call one of them default – they are just accepted to be selected by phonology (I have to note that according to Chomsky they are not even selected but are only *one* underlying suffix which is modified by pronounciation rules). Still, an experiment is set up to strengthen the position that both Dutch plural noun suffixes are equally regular.

2.2 A wug-like experiment to check productivity

To check productivity of both suffixes, the authors have created a set of 80 fantasy words which cover nine possible criteria about which suffix to select. All words are built to be possible Dutch nouns, some of them selecting a certain plural suffix by phonology, some only preferring a certain plural suffix, and some influenced by morphological effects (such as *bestroeting* as being analyzed *bestroet-ing*). Subjects where asked to write down plural forms for the fantasy words, and the ratio of *-en* and *-s* suffixes they used was then analyzed.

The outcome was as expected: Where the phonology selected a certain suffix, allmost all subjects used that suffix for the fantasy words. In the less restrictive cases, both suffixes were used: For words like *kna*, about 4 out of 5 cases got an *-s*, leaving still almost 19 percent cases for the *-en* suffix. For three other constructions, the distribution was the other way round, still giving at least about one quarter of the cases to the dispreferred suffix. So it can be clearly seen that *both* suffixes are productive and fully regular.

3 Surface Frequency Effects with Dutch verbs

As we have seen, it is problematic to talk about a single default inflection for Dutch noun plurals. So the authors prefer to continue with experiments in a less debatable area: Verbal inflection. In the analyzed cases, there is only one very frequent affix to realize a given inflection, so the question of default is avoided. Still, they manage to find Surface Frequency Effects, giving evidence for some storage of inflected forms in parallel to the clearly rule-based inflection in the analyzed group of regularly inflected words.

3.1 Perfect Participle

The Dutch perfect perticiple is formed with the *ge- -D* (D can be realized as *-d* or *-t*, the authors have only used *-d* forms according to ortography). The experiment uses two groups of participles, which only differ in their average *surface* frequency, but are matched in base frequency, lenght and family size. Thus, if an effect is detected, it should be one of inflected regular forms being stored as a whole even though a rule can handle this kind of inflection very well. The idea is that for frequent forms, storage of the inflected form leads to even better performance than running the word from the base form trough the rule to the inflected form or back.

The experiment was done as a *visual lexical decision* task: The subjects we-

re shown strings on a screen (first, a fixation mark was displayed, then, after a fixed time, the string, for a fixed time) and had to decide as quickly as possible but still accurate whether the shown string was (an inflected form of) a real Dutch word.

The results did show a clear Surface Frequency Effect: The more common forms were recognized faster and more reliable than the rare forms, which shows – according to the authors – that their use is supported by a memory mechanism faster than the rule, at least given a high surface frequency.

The authors did some further analysis of the data and found out that there was a Base Frequency Effect in the other direction: Their explanation is that a common base (lexeme) will more readily be fed into the rule system, which will increase competition of the rule based system with storage system that may have the inflected form on offer as well, thus slowing down the decision process. However, one may also expect an effect in a way that the slower a base is retrieved, the more it has to suffer from the competing retrieval of the inflected form. I take it that it depends considerably on the structure of the system which processes can compete at certain points: It is well possible that the base lexicon lookup is independend of the competition for having the right inflected form (which is on the other end of the rule processing pipeline, one could say).

Yet another explanation settles on the speed of the rule/base lexicon system: A segmentation of, for example *ge-wandel-d* may be available quickly and actually confuse and slow down the parsing process because the segmentation contains the misleading partial parse *ge-wandel*. This would happen especially in the case of frequent base lexemes. Where the segmentation cannot be confused with a *ge-base* combination, the authors found some positive effect in a way that frequent bases made processing easier in this case, which gives further support to this explanation.

In conclusion, the authors argue that Surface Frequency Effects are not limited to irregular forms. As the data both shows parsing effects and storage effects, both processes seem to participate in handling the regular inflection of the Dutch perfect participle.

Next, the authors extend their study from inherently inflected forms such as noun plurals (where the inflection is part of the semantics, one could) and less clear cases such as perfect participles to verb plurals: Verb plurals are a kind of contextual inflection, as they are controlled by agreement in the local context and thus less likely to be found stored as full inflected forms in some storage area in the brain.

This is to be seen in context to other experiments by Baayen et al. and *Bertram et al.* (who has also done some research on Finnish), where no Surface Frequency Effects were found for the past tense suffixes *-te* and *-en*. Thus the effect just found may be due to some other effect such as the misleading parse mentioned or to the kind of word formation used. The next experiment will be about the past tense plural inflection, as this is a contextual and thus arguably more prototypically kind of inflection.

3.2 Past Tense Plural

This experiment had a very similar setup to the experiment just discussed, but this time, the past tense plural forms ending in *-den* were the object

of examination. Again, two sets of verb forms were selected, both with similar base frequency, length and family size, but different surface frequencies. The experiment was done in the same run as the next experiment, doing both at the same time with the same group of subjects.

Again, a reliable Surface Frequency Effect could be detected. The response times and error rate were lower for forms with higher surface frequency. This gives *again* evidence for storage interfering with regular inflection, and this time the inflection is even more prototypically regular, as there are no competing suffixes (as with the plurals) and the inflection is contextual rather than inherent. The effect is smaller, but still reliable. In this experiment, no clearly independend Base Frequency Effect was found.

The authors compare their results to the results of Bertram et al., who did not observe Surface Frequency Effects for the singular past tense suffix *-te*. Baayen et al. argue that the reason is the less sensitive experimental setup, because Bertram et al. have used a much smaller frequency contrast.

3.3 Present Participle

Last, the authors did an experiment on the present participle, which is in itself something used not very often – so the general surface frequency is low. The *-end* suffix (as in *wandelend*) is fully regular and productive and has no rival alternative suffixes: It is a default.

Using the same basic setup for the experiment, but with a lower overall frequency, the Baayen et al. were still able to observe a reliable Surface Frequency Effect similar to the one in the Past Tense Plural experiment. So the effect is quite robust, as this last experiment involved a smaller frequency contrast, low overall frequency and a very clear case of regular inflection.

There was also a Base Frequency Effect, where more frequent bases were correlated with faster response times, although there was no reliable correlation between surface frequency and base frequency. The explanation of Baayen et al. is that for those generally low frequency words both the parsing route and the memory retrieval contribute to finding the full inflected form at the same time.

4 Results and Discussion

4.1 General Things

Combining all results, the authors present strong evidence that not regularity (as assumed by other dual route models such as the one advocated by Pinker) but frequency is the main factor for determining the weight of storage versus calculation. Storage is not only limited to cases inexplicable by rules, but can extend to any case where a high surface frequency promises some gain over using rules alone by storing inflected forms. This contrasts to the viewpoints of Marcus et al. and Clahsen et al., who are relying on the default (in the sense of prototypically regular) status of a suffix to decide it has to be handled solely by rules.

However, Baayen et al. do not claim frequency to be the only factor: Rather, they assume the balance of storage and computation being the result of a complex process, involving frequency, complexity of the involved calculations, difficulty of storage, and so on.

So they have weighted costs for rule application and memory space and access and the frequency being a main factor in determining the gain that is rewarded for decreasing the cost of certain instances of inflection.

In the text on which the given summary is based, the results part contains a considerable amount of repeating and summarizing what has been first announced in the introduction and then worked out to quite some detail in the main part. Given that this already is a summary, I do not summarize their summary again...

I also avoid repeating their discussion of when effects are observable and when not (e.g. progressive demasking being more sensitive to surface effects than visual lexical decision, because the latter also involves processing of meaning to some extent) and possible levels of representation in storage (because their notion is a bit fuzzy and hard to understand).

4.2 Predictions on the Balance of Systems

The authors have shown that several cases of regular and even default inflection do show Surface Frequency Effects, which conflicts with the assumption that the handling of regular defaults in particluar and rules in general is mutually exclusive to memorizing forms. The findings conflict with some dual route models where only irregulars are memorized but the rules are always active (but blocked if a memorized form is found) in a way. However, there are some common points with those dual route models: Given the assumptions of Baayen et al., both systems are always concerned with analysis or production of inflected forms, and unless competition effects arise, overall performance is increased by combining the best performance of both. So there is nothing to say against memorizing a frequent regular form, as it reduces the need to run the rules all the way through frequently while not causing too high storage costs (as high token frequency of a form usually coincides with the form being one of a few frequent ones).

As is added as a new point in the results and discussion sector, doing a dot plot of reaction time versus log full form (surface) frequency shows a large variance, but further analysis does support a linear dependency over a wide range of frequencies. The reaction time seems to reach a maximum below a very low frequency, and even this can be due to sparse data problems. So – as opposed to *Allegre* and Gordon – there seems to be no or at least a very low threshold below which memory ceases being used to speed up parsing.

So one can no longer ask what is stored and what not, but one has to ask what determines the balance of storage and computation. Both the radical connectionist way of storing everything and the radical classical way of allowing only irregulars into some memory list seem to be too strict with their claims. The offered factors influencing the balance involve cost of storage and computation, frequency, but also (as I will explain below) the modality and other factors.

4.3 The Costs of Storage and Computation

For the frequency, both the base frequency and the surface frequency have to be taken into account: The surface frequency as a predictor of how often the inflected form is seen, and the base frequency to how often a word base

is one ones mind in general. The base frequency has more influence on ease of handling by the rule system, while the surface frequency is more of a storage shortcut to reduce the need of using the rules for frequent forms. As the rule system runs in both directions, the interferences can be of a complex kind, including competition effects and the full-form storage hinting or priming some processes of the rule system or the base lexicon.

The costs of storage are not easily calculated, but it can be said that the total storage available is quite huge, giving the possibility to store many inflected forms for languages with simple morphology such as English. Still, languages with rich inflection such as Turkish or Finnish would cause a big load on the storage system if one was to memorize too many inflected forms, thus slowing down the access. Research by *Niemi et al.* on Finnish fits that idea by not showing Surface Frequency Effects in Finnish word formation.

The costs of calculation depend on the kind of calculation to be done: sometimes using memory rather than parsing can speed up processing, but this depends on several factors as stated above. *Schreuder* and Baayen suggest a parallel dual route model, where both parsing and storage work in parallel and the first route to finish will to a large extend determine the result and the response time of the system. In this model, as the timing gets too similar, competition between both routes will arise and slow down processing.

Baayen et al. point out that it is important to distinguish between language production and comprehension: In case of recognition, there is no need for cases of irregular inflection blocking some regular default, because overregularization effects are irrelevant as the correct inflected form is already part of the input. This does not, however, mean that there are separate devices for handling regular and irregular inflection.

They also take this distinction to be important for the minority default argument of Marcus on the German noun plural: It is counterintuitive that a *-s* default would only constitute seven percent of the noun types and two percent of the noun tokens, and storage of all the regular *-en* and *-e* cases among other less regular cases seems also to be implausible. But as it is felt to be much easier to understand than to produce German noun plurals, a close explanation would be the use of more parsing in comprehension, while production is – although allegedly also using rules much more than Marcus assumes – complicated by troubles selecting the right one of several rules.

Still, Baayen et al. hold that storage may be in use to some degree virtually everywhere. They argue that the answer to the question of the balance of storage and computation has to be much more complex than some well-known mottoes such as *store what rules cannot capture* on one hand or *rules are only a fallback for memory failures* on the other hand.