

# HORACE

## —an artificial columnist

Marcus Uneson, Jan 2003

### Abstract

*After a brief outlook on the field of random text generation, in particular on Andrew Bulhak's Postmodernism generator, the present paper describes a program for generation of random, meaningless but grammatically correct text in Swedish. The program, named Horace, is intended to simulate the abstract reasoning of (some) literary columnists.*

*Horace is written in Prolog using the DCG formalism. It handles agreement and permits weights to be assigned to competing rules. A first version can be tested at Horace [www](#) (embedded in a Perl CGI for [www](#) access). The paper is concluded with suggestions of various experimental, application-specific extensions to the program.*

### Introduction

#### Automatic text generation

Automatic generation of text from some underlying, formal semantic representation is an important research field. Generally, the point of automatic text generation (or “natural language generation”) is to adequately render a system- and application-specific machine representation (which is very opaque to human beings) in natural language (which is, at least sometimes, immediately and effortlessly accessible).

For restricted domains good results have been attained. On a more general level, however, most things remain to be done. Automatic generation of text in the sense of transforming a formal semantic representation into a grammatical and (above all) coherent text is a very demanding task (for a first introduction, see for instance Gal et al 1991). Nevertheless, the field is most central to improvements on the interface between man and machine. Research in the field is quite vivid; for instance, the 2nd International Natural Language Generation Conference was held 2002 and there have been several international workshops sponsored by the Association for Computational Linguistics (ACL [www](#)). There is also an ACL Special Interest Group on Text Generation, SIGGEN (Siggen [www](#)).

#### Random texts

The topic of the present paper, however, is another: generation of *random* texts. This task is easier by several orders of magnitude. For such completely meaningless texts, there is no semantic representation at all to be conveyed. Several difficult models (of semantics, pragmatics, world knowledge, and discourse structure, among others) are thereby made superfluous.

Random text generation may seem as pointless as the texts themselves.

Admittedly, the applications from a practical point of view are few: Turing tests, tests of a grammar under construction, perhaps generation of sentences for language education, when the teacher's imagination is exhausted. The AI and computational linguistics research communities seem to have largely lost interest in the field (perhaps after the successful but now terminated attempts with the semantically clueless ELIZA and PARRY), turning the attention to automatic text generation in the sense described in previous section.

However, the main *raison d'être* of random texts is diversion, and diversion will probably never become out of date. For a more general audience, simulation of texts produced by humans in different genres continue to attract interest. On the web one may find for instance randomly generated poetry (Kurzweil [www](#), Zawinski [www](#)), buzz-word spoof commercials (Lee [www](#)), pulp fiction covers (Romance writer [www](#)), pseudo-philosophical ranting (Kant [www](#)), or postmodern discourse (Postmodernism [www](#)). See the link collections Toolworx [www](#) and Charabia [www](#) for more links.

## On the present program

The current program, named Horace, is certainly not intended for anything but amusement. It attempts to imitate the discourse of certain cultural columnists of the Swedish intelligentsia. As pointed out by Bulhak 1996, the vocabulary of such writers is often abstract, dense, and replete with jargon; additionally, the disciplines they comment (art criticism, philosophy, cultural theory, among others) are inherently subjective, with reasoning based on analogies, comparisons and references to text-external sources, rather than on logic. Thus, this kind of writing should be easier than the average to simulate. The texts generated by computers certainly may appear incoherent, incomprehensible, and difficult to follow; but then again, for this specific genre, so may those generated by humans.

It is by pure coincidence that Horace bears the same name as Horace Engdahl, permanent secretary of the Swedish Academy and notorious for his esoteric literary reviews.

## Strategies for random text generation

The most popular approaches for random text generation may roughly be divided into statistical modelling and explicit grammars. They are briefly commented on below.

### Statistical modelling

Statistical modelling is stable, easy to implement and covers inherently collocation (at least for immediate neighbouring words, such as *carry through*, *state senate*), which probably is an important cue for naturalness. A common method is to construct a training corpus; to divide it into units (most often words, sometimes individual characters); and to construct a probability table (a *language model*) wherein a probability is assigned to each unit, given some preceding sequence of a certain length (the *degree* of the model; for words, typically three or four).

Such a model encodes very little or no linguistic knowledge—the text is treated as a sequence of arbitrary symbols from a given alphabet and would need very few modifications to analyse, say, amino acid or DNA sequences. While this is perhaps a weakness only from a linguist's point of view, there are more substantial drawbacks as well. Thus, a training corpus

must be prepared—for some purposes, this task may amount to downloading a collection of texts from www, but it may also include very laborious and tedious work (as in this case—corpora of newspaper text exist, but a large enough corpus of literary criticism in Swedish, preferably by one or two particularly abstruse authors, is not easily found).

Furthermore, the statistical model has structural deficiencies: for one thing, it does not easily handle dependencies reaching outside the window width set by the model's degree. This may be less obvious in English than in most other languages, for which agreement is more important. In modern standard Swedish, verbal agreement is even simpler than in English (i.e. non-existent); the rules for NP agreement, however, are quite complex (for instance, different adjectival agreement for definite and indefinite and for attributive and predicative position). Some simple examples of agreement outside the span (for a 4-gram model) are given below:

[1a] de otroligt stora, gröna ängarna  
'the incredibly large, green meadows'

[1b] \*de otroligt stora, gröna ängen

[2a] den otroligt stora, gröna ängen  
'the incredibly large, green meadow'

[2b] \*den otroligt stora, gröna ängarna

Given the sequence {otroligt, stora, gröna}, there is no way in a 4-gram model to correctly choose between the singular [1a] and the plural [2a] by pure statistics.

## Explicit formal grammars

Grammatical correctness (on which agreement is but one aspect) seems to be *sine qua non* for successful simulation. Whereas many writers may get away with flawed reasoning, blurred semantics or general incoherence (I suppose most of us do, occasionally), such failures are not as immediately obvious to a casual reader as are incorrect endings, erroneous number of verb arguments, or (for languages which so require) failure to include subject and/or finite verb in each clause. To model long and complicated but immaculately grammatical (in the Chomskyan sense) sentences, an explicit formal grammar is called for.

It should be noted that writing a grammar for generating text is far easier than writing one for parsing (on a general level, that is). In parsing, it is necessary to foresee and formalize the possible constructions of a language, which is a tremendous challenge; in generation, by contrast, it is enough to specify the constructions wanted in the generated text. Grammatically correct but highly unlikely constructions (such as nested relative clauses, or recursively called prepositional phrases beyond a depth of, say, three) may be discarded with no immediate drawbacks. Furthermore, if the output isn't very long (as in the current case), some perfectly normal constructions that happen to be difficult to formalize or implement efficiently (in Swedish, for instance, movements, particle verbs, prepositional complements), can be left out without losing much naturalness. Repetitions of a certain grammatical structure is by no means as conspicuous to a human reader as repetitions of an unusual lexeme, and we note what's there rather than what's not.

*Explicit grammars in use: The postmodernism generator*

Most of the examples of web sites featuring computer-written texts cited above appear to use some kind of statistical modelling; those that do not are mostly extremely simple (offering perhaps a three-word phrase with a randomly chosen verb, followed by a randomly chosen adjective, followed by a randomly chosen noun). One notable exception is the aforementioned “postmodernism generator” (Postmodernism www), the most ambitious attempt at simulating genre-specific text I have come across.

As mentioned before, random text generation has generally not attracted much interest from the research community in the latest years. The postmodernism generator is an exception in this sense as well: it is described in the paper, “On the Simulation of Postmodernism and Mental Debility using Recursive Transition Networks” by Andrew C. Bulhak (1996). The paper presents briefly a system called “The Dada engine”, which accepts as input a script defining a set of rules in form of recursive transition networks (RTNs) in an especially devised format called pb (slightly reminding of the Backus-Naur form). Bulhak has provided some extensions to the basic RTN approach; in particular, the rules may take parameters, which permits lambda abstraction. The engine traverses the RTNs, choosing rules at random, and outputs strings.

A particularly successful set of RTNs simulating postmodern discourse in the style of a journal article is also described in the paper (Postmodernism www offers an online version). A few excerpts from the randomized postmodernist article “Realities of Stasis: Subsemiotic materialism and Foucaultist power relations” (included as a sample in Bulhak’s paper) are given below:

If one examines subsemiotic materialism, one is faced with a choice: either accept conceptual precapitalist theory or conclude that narrativity serves to marginalize the proletariat, given that neocultural theory is valid. Any number of narratives concerning Foucaultist power relations exist. Subsemiotic materialism implies that sexuality has objective value.

(...)

Lyotard promotes the use of Marxist socialism to attack the status quo. Foucaultist power relations suggests that culture is capable of deconstruction. However, several deconstructions concerning subsemiotic materialism exist.

(...)

In a sense, a number of narratives concerning subsemiotic materialism exist.

(...)

If one examines textual capitalism, one is faced with a choice: either accept Foucaultist power relations or conclude that the goal of the reader is deconstruction.

(...)

If one examines Foucaultist power relations, one is faced with a choice: either reject textual capitalism or conclude that truth is capable of intention, given that art is equal to narrativity.

The approach does have its limitations. Judging from the sample article document cited above, Bulhak has attacked the problem from a computer scientist’s point of view (rather than a linguist’s). Terminals may have any length, with little consideration of linguistic relevance, so some parts remind of filled-in templates. Furthermore, the RTNs do not easily handle agreement (not even with Bulhak’s extensions), which makes the engine difficult to rewrite for other natural languages. (At times this is noticeable even in English. All verbs output by the postmodernism generator are in third person singular, present tense. So are almost all subjects as well, but occasional exceptions

are not considered—note the ungrammatical “Foucaultist power relations suggests” above). Another weakness is the primitive way of assigning individual weights to rules; currently, any rule with proper head is as likely to be chosen as another, and doubling its probability is done by including it twice in the script. This is not a very sophisticated approach, and not something you would like to use to assign weights {10000, 30, 1} to {rule1, rule2, rule3}.

Still, the overall impression of a paper produced by the postmodernism generator is quite convincing. The repertoire of constructions is large enough not to make the repeated structures too obvious (the quotes above are chosen from a three-page document), and the output even includes made-up quotes and references. The program has attracted a fair amount of attention. In October 2002, it participated in an art exhibition called Electrohype 2002 (Malmö, Sweden; Electrohype [www](http://www.electrohype.com)).

## Horace

### Issues of formal grammars for Swedish

The modelling of Swedish grammar for random text generation presents some difficulties which are absent or at least less cumbersome in English. In particular, they concern agreement constraints and word order. Some other difficulties, less ubiquitous (e.g. particle verbs, movements, prepositional objects), have simply been outlawed—see under “Restrictions” below.

#### *Agreement*

Agreement constraints in Swedish concern in particular the nominal phrase (see also examples [1-2]). Adjectives can be regarded as having inflections for number, gender, species (definite/indefinite; compare [3a] to [3b], and also [3c] to [3d]), and position (attributive/predicative; compare [3b] to [3d]).

[3a] en färglös idé/ett färglöst minne/färglösa idéer  
'a colourless idea/a colourless memory/colourless ideas' (attributive position, indefinite)

[3b] den färglösa idén/det färglösa minnet/de färglösa idéerna  
'the colourless idea/the colourless memory/the colourless ideas' (attributive position, definite)

[3c] en idé är färglös/ett minne är färglöst/idéer är färglösa  
'an idea is colourless/a memory is colourless/ideas are colourless' (predicative position, indefinite)

[3d] idén är färglös/ minnet är färglöst/idéerna är färglösa  
'the idea is colourless/the memory is colourless/ideas are colourless' (predicative position, definite)

A fifth conceivable category is natural gender. In written Swedish, the adjective ending is often -e for masculine persons in definite singular attributive; compare [4a] and [4b].

[4a] den store hjälten  
'the great hero' (masculine)

[4b] den stora hjältinnan  
'the great heroine' (feminine)

However, for many writers, this is not a compulsory distinction, and it is currently unimplemented in Horace.

### *Word order*

Swedish is a Germanic language and, like several others of those, it has strict rules for where to place the finite verb. The constituent order is often described by means of the Danish linguist Paul Diderichsen’s position schema (originally for Danish), as in Figure 1:

Foundation	Nexus field			Content field		
	V1	N1	A1	V2	N2	A2
	finite verb	subject (when not in foundation)	clausal adverbials, short adverbials	infinite verb/s, verb particles	object/s, predicative	time, place, manner etc adverbials

Figure 1a. Diderichsen’s position schema, main clause

Subjunction field	Nexus field			Content field		
	N1	A1	V1	V2	N2	A2
	subject	clausal adverbials, short adverbials	finite verb	infinite verb/s, verb particles	object/s, predicative	time, place, manner etc adverbials

Figure 1b. Diderichsen’s position schema, subordinate clause

Basically, there are two different schemata, one for main clauses and one for subordinate clauses. In virtually all declarative main clauses, the verb goes into the second slot (Fig 1a), and the first slot (the foundation, Diderichsen’s *Fundament*) can and must be filled with exactly one of the other constituents<sup>1</sup>. The entire phenomenon is known as “V2”, which may be interpreted as “verb in second slot”. The by far most common content of the foundation is either subject (N1) or clausal/short adverbials (A1). The constituent order of subordinate clauses is more rigid.

## Implementation

### *Requirement specification*

Horace is expected to generate a text consisting of grammatically immaculate sentences composed from an entirely abstract vocabulary. A later step is to include a rudimentary simulation of discourse structure (such as rhetorical markers and/or references to other fictive writers).

### *Formalism*

Horace is written in Prolog, using the DCG formalism (Pereira and Warren, 1980). DCG was initially chosen for Horace as a convenient way of handling agreement. However, with one argument for probability, one for unique ID (more on these below); perhaps three or four for morphological and agreement categories; and two for difference lists (when treating the implicit lists of DCG rules in ordinary prolog code), the argument lists may appear quite cluttered. For rules involving many constituents, such as instance ditransitive verbs with optional adverbials, the notation does become cumbersome.

Additionally, DCG does not handle the V2-mechanism in some obvious way. The program may one day be rewritten in some other formalism, such as Constraint Grammar.

### *Allowed constructions*

For the purposes of Horace, it is enough to provide a grammar which is extensive enough to generate some 40 lines of text without too obviously repeating grammatical constructions. As pointed out, human language perception is not very easy to offend in this sense—if the vocabulary is varied, a readable article may be generated with rather few rules. Quite a few constructions may thus be simplified or disregarded. The V2 phenomenon can be reduced to the two most common cases: subject or adverbial in the foundation. Wh-movements may be disregarded, as may particle verbs, prepositional complements, any recursive categories, ellipses, extragrammatical utterances etc

### *Non-terminals*

It is in fact easier to state what the grammar does cater for, than what has been disregarded. Table 1 gives an overview over the non-terminals currently used, with an informal use of regular expression modifiers (`{}`|`?`) for quantification and disjunction, and DCG-style square brackets (`[]`) for terminals. All DCG arguments (for weights, rule ID, morphological categories, agreement etc) have been left out for clarity, as well as methods for choosing terminals and rewrite rules at random (see below). Variations for main and subordinate clause have also been left out; they include active vs passive clauses, finite vs compound verb forms, optional adverbial phrases, and (for main clauses) subjects vs clausal adverbials in the foundation.

**Table 1.**

Non-terminals of Horace.

heading	[Art] [Adj] [N]	heading
critic	paragraph{3}	critic
paragraph	s{6,9}	paragraph
s	mcl   mcl scl   scl mcl   mcl [Conj] mcl	sentence
mcl	cl_advl [Aux] np (vit   vmt np   vdt np np   vkp np   vkp adjp)	main clause (only one type shown)
scl	subj np cl_advl (vit   vmt np   vdt np np   vkp np   vkp adjp)	subordinate clause (only one type shown)
np	n_grp (pp)?	nominal phrase
np	gen_attr (adjp)? [N]	
subst_adjp	[Art] (advp)? subst_adj	substantivized adjectival phrase
gen_attr	n_grp	genitive attribute
n_grp	([Art] (adjp)? [N])	noun_group
n_grp	subst_adjp	
adjp	(advp)? [Adj]	adjectival phrase
pp	[P], n_grp	prepositional phrase
agent	[Agent_marker] np	agent
advp	[Adv]	adverbial phrase
cl_advl	scl   advp   [Cl_adv]	clausal adverbial

### *Terminals*

Terminals are handled by an abstraction layer which hides details like inflection class (declination or conjugation), implemented as follows:

The available vocabulary is read in from a separate file (lexicon.txt) at start-up. As an example, let's consider a small sample from noun declination 3:

```

%%% in file 'lexicon.txt'

noun([decl3, utr],
      [intighet, poststrukturalitet, kausalitet]).

%%% directive in file 'horace.dcg' (last in file)

:-
    consult('lexicon.txt'),
    abolish(nouns_/1),
    nouns(N),
    assert(nouns_(N)).

%%% in file 'horace.dcg'

nouns(N) :-
    setof(Decl, List^noun(Decl, List), Decls),
    list_all_infl(noun, Decls, N).

decline_noun(Nlemma, sg, indef, decl3, N) :- ccat(Nlemma, '', N).
decline_noun(Nlemma, sg, def, decl3, N) :- ccat(Nlemma, 'en', N).
decline_noun(Nlemma, pl, indef, decl3, N) :- ccat(Nlemma, 'er', N).
decline_noun(Nlemma, pl, def, decl3, N) :- ccat(Nlemma, 'erna', N).

case_infect_noun(N, nom, N).
case_infect_noun(Nnom, gen, N) :- ccat(Nnom, 's', N).

noun_agreement(pl, _, pl).
noun_agreement(Num, Gen, Gen) :-
    Num \= pl.

%%% general pos-list-building predicates

%builds a list of form pos
list_all_infl(_, [], []).
list_all_infl(POS, [Infl|InflRest], Out) :-
    list_one_infl(POS, Infl, Out1),
    list_all_infl(POS, InflRest, OutRest),
    append(Out1, OutRest, Out).

list_one_infl(POS, Infl, Out) :-
    Term =.. [POS, Infl, Wordlist],
    clause(Term, true),
    make_list(Wordlist, Infl, Out).

make_list([], _, []).
make_list([Word1|WordRest], Infl, [[Word1, Infl]|Rest]) :-
    make_list(WordRest, Infl, Rest).

%returns an Element randomly chosen from List
randomize(Element, List) :-
    length(List, Max),
    Random is random(Max),
    length(Left, Random),
    append(Left, [Element|_], List).

%concatenates atoms X and Y to atom Z
ccat(X, Y, Z) :-
    atom(X), atom(Y), var(Z),
    name(X, XL), name(Y, YL),
    append(XL, YL, ZL),
    name(Z, ZL).

```

After these preparations, a predicate `find_declined_noun` may be defined as below, which returns a random noun in specified number, species, and case; it also returns the `GenNum` agreement marker for further processing. This marker



has one of three string values: ‘ntr’ for the neuter and ‘utr’ for the reale (also known as “common gender”, “non-neuter”, or “uter”) in singular, and ‘pl’ in the plural.

```
% (+Num, +Spec, +Case, -GenNum, -N)
find_declined_noun(Num, Spec, Case, GenNum, N):-
    nouns_(Ns),
    randomize([NLemma, [Decl, Gen]], Ns),
    decline_noun(NLemma, Num, Spec, Decl, Nnom),
    case_inflect_noun(Nnom, Case, N),
    noun_agreement(Num, Gen, GenNum).
```

Most other parts-of-speech are treated similarly. Verbs with different arities (intransitives, monotransitives, ditransitives) are treated as belonging to separate parts-of-speech. For the open classes, terminals may thus be specified using this type of “find\_one\_random\_X\_with\_this\_inflection” predicate.

In some cases, generation of inflected forms by simple concatenation of atoms (as in `decline_noun/5` above) isn’t enough. An unstressed ending vowel present in the lemma is generally not part of the stem and therefore lost in inflected forms (pojke-ar > pojkar, blomma-or > blommor). Another common mechanism is the deletion of a stem-final dental stop (/d/, /t/) before inflecting adjectives and participles for the neuter (skadad-t > skadat; immanent - t > immanent).

While Horace does handle these two standard cases, several computational challenges from a morphological point of view has been disregarded. Some of them (e.g., umlaut) aren’t easy to handle with simple string concatenation. However, the intended, abstract vocabulary of Horace is almost entirely of Latin or Greek origin (for the open classes). Such late loans are generally quite regular morphologically. For instance, in Swedish, almost all verbs of Latin origin end up in the very regular first conjugation (*abstrahera*, -r; -de, -t, -s, -s, -des, -ts, -nde, -d, -t ‘to abstract’); and many adjectives end in -isk (*logisk*, *ironisk* ‘logical, ironical’ etc)

### *Lexicon*

The lexicon is constructed by hand, to keep the abstraction level as high as possible. As pointed out, the late loans aimed at are morphologically very regular; thus, a few searches on ‘-isk’ and ‘-era’ in a larger Swedish corpus (e.g. Språkbanken [www](http://www.sprakbanken.se)) provided valuable help.

## Weighting

In contrast to the postmodernism generator, Horace does permit the assignment of user-specified weights to each rule. For competing rules, individual integer weights are assigned as the first argument, and a unique rule ID as the second. The weights could conceivably be extracted automatically from a training corpus yet to be built; currently, however, they are somewhat arbitrarily set by hand. The predicate `find_idx(Q, IDX)` then

1. takes Q (a non-terminal to be rewritten, perhaps with some arguments specified) as input;
2. searches through all rules which are applicable for the particular combination of constraint arguments;
3. sums weights of the applicable rules (first argument);
4. picks a random number integer i,  $0 < i < \text{sum\_weights}$ ;
5. maps i to an ID considering the weights;
6. returns that ID.

```

find_idx(Q, IDX) :-
    findall([ID, Wt],
        (
            clause(Q, _),
            arg(1, Q, ID),
            arg(2, Q, Wt)
        ), IdWts),
    sum_wts(IdWts, 0, TotIdWt, IdAccWts),
    N is random(TotIdWt),
    id(N, IdAccWts, IDX).

id(N, [_ , AccWt | RestIdAccWts], IDX) :-
    AccWt < N,
    id(N, RestIdAccWts, IDX).

id(N, [[IDX, AccWt] | _], IDX) :-
    AccWt >= N,
    !.

sum_wts([], Acc, Acc, []).
sum_wts([[ID, Wt] | Rest], Acc, TotIdWt, [[ID, NewAcc] | RestIdWts]) :-
    NewAcc is Acc + Wt,
    sum_wts(Rest, NewAcc, TotIdWt, RestIdWts).

```

## DCG extensions

Horace is likely to be expanded according to the feedback it elicitates. The following are suggestions of future extensions to the DCG used in the current program. They are all experimental at most; some are very sketchy at the time of writing and may indeed never be implemented. On the other hand, there may be others instead.

The extensions typically work with entire phrases and should not be regarded as grammatical statements about an entire genre; rather, they are implementationally cheap, application-specific enhancements of the impression of an erudite columnist.

### *Simulation of discourse (experimental)*

The general impression of reasoning is highly enhanced by the presence of (some simulation of) organized discourse. This is (in Swedish, as in most Western languages) most naturally attained by using discourse keywords, giving the impression of collecting pros and contras, arguments and objections, before arriving at a conclusion. Like before, the task is far easier in generation than in parsing, since for a reasonably short text, the human reader will concentrate on what's present in the text, rather than what is not.

A way of simulating a trace of reasoning is to add a few categories, say <statement>, <concession>, and <conclusion> for rhetorical structure organizers. A pseudo-code style attempt (with little attention paid to word order and clause borders) could go along the following lines:

```

rhetoric_segment --> reasoning, conclusion
reasoning --> statement, concession
statement --> mcl

concession --> concession_phrase, scl
concession --> concession_marker, mcl
concession_phrase --> [det, är], concession_marker [sant, att]
%while it is true that...
concession_phrase --> [det, är], concession_marker [så , att]

concession_marker --> [i, och, för, sig]; [visserligen]

```

```

conclusion --> conclusion_phrase, scl
conclusion --> conclusion_marker, mcl
conclusion_phrase --> [det, är], conclusion_marker, [klart, att]
%it is clear that...
conclusion_phrase --> [det, är], conclusion_marker, [tydligt, att]
conclusion_phrase --> [det, är], conclusion_marker, [uppenbart, att]
conclusion_marker --> [följaktligen]; [sålunda]; [alltså]; [således]

```

### *Keywords and phrases (not yet implemented)*

A conspicuous property of the output of the postmodernism generator is the high frequency of recurring short phrases, typically NP:s such as N +N, Adj + N, or Adj + N + N (in the essay quoted above, for instance ‘subsemiotic materialism’, ‘neocultural theory’, ‘Foucaultist power relations’). It is interesting to note that these repetitions, while probably unnatural in a novel or most other genres, in the pseudo-scientific jargon of the thesis quoted rather contribute to a vague impression of terminological consistence.

Horace aims rather at a newspaper columnist writing style, and repetitions of this type are somewhat less likely to occur. However, there may certainly be a point in having a few named concepts recurring. Most obviously, for a review in literature or art, the name of the artist and the work reviewed should be mentioned now and again.

In the Postmodernism generator, such fixed phrases are hard-coded, recurring in essay after essay. Another approach is to generate a few phrases at run-time and to have them repeated with a certain probability (for the artist, perhaps in subject position only).

A related question is that of pronominalization. An ordinary text with no pronominalizations appear highly unnatural and over-specified. However, for the highly specific impression aimed at here, with loads of piled-up abstractions, they are less crucial. Still, whenever a person is mentioned (most notably the artist), it is natural to use a pronoun within the following clause or two.

### *Parenthetic clarifications (experimental)*

Horace could explain or expand difficult concepts to the reader by including a quoted clarification in parentheses:

```

koreografien i det voyeuristiska blir extasens poststrukturalism ("det suberotiska
elementet").
'the choreography of the voyeuristic turns into the poststructuralism of ecstasy
("the suberotic element")'

```

The explanation pertains to the same syntactic category as the phrase explained and thus needs little extra modelling. In principle, any phrase type could be expanded in this way; however, the idea seems to work most efficiently with NP:s, in which case the added explanation somewhat reminds of a definition.

### *Quotes from other great thinkers (not yet implemented)*

Horace could easily find support for its claims by enclosing arbitrary clauses in double quotes and ascribing them to some authority on the subject, perhaps:

```

Som Derrida påpekar: "intighetens sant labyrintiska kontrapunktik delegerar
konstruktivismen i det enigmatiska"
'As Derrida points out: "the truly labyrinthic counterpoint of nullity delegates
the constructivism of the enigmatic"'

```

### *Neologisms (experimental)*

Any abstract noun or adjective (*transcendental, intighet, modernistisk*

‘transcendental, nullity, modernistic’) may be combined at random with a derivational prefix, such as for instance *meta-*, *neo-*, *hyper-*, *hypo-*, *sub-*, *pseudo-*, *kvasi-* ‘meta-, neo-, hyper-, hypo-, sub-, pseudo-, quasi-’. The result is an even more abstract neologism (*kvasiintighet*, *metamodernistisk*, *subtranscendental* ‘quasi-nullity, meta-modernistic, subtranscendental’). A given prefix should be used no more than once per article.

## Web presentation

Horace is currently available at Horace www (not including the experimental features). Input (name and sex of artist; sex is at the time of writing not used but will be needed for planned extensions like pronominalization) is validated by a cgi script in Perl and then forwarded to the prolog DCG. The simple, xml-like output of the DCG is returned to the cgi script in Perl, which pours it into an HTML template in newspaper column style. The fake article also carries an illustration (an abstract painting randomly chosen at runtime; the painting is currently chosen from a collection by the abstract expressionist Jackson Pollack).

## Sample output:

as for version of Dec 10, 2002 (text only).

Det kosmetiska i det illusoriska debatteras av en ekosofi, eftersom de geometriska surrealismernas retorik har förlänat det dekonstruktivistiska de obevekligt akademiska anomiernas eugeni. Trots att det disharmoniskas asymmetri har deklarerat fanatismen blir det efterhängset dialogiska musikantiskt. Symbiotiker har applåderat, eftersom surrealismens postfotografism förlänar morfologismer i de kaleidoskopiska allegorierna sekterismens semiotik. Naturligtvis abstraherar det postmodernistiska i det oundvikligt elegiska kameleontismer. De pandemiska hermeneutismerna blir inte symfoniska, emedan aristokratiens obevekligt nupsykologiska ekvilibrism förlänar de holistiska gnosticismerna det övergripande atlantiskas plastisk. Det sant konstruktivistiska i pekoralismen har aktualiserats, emedan arkaismer har förlänat letargismens hermesi allegorien. Trots att neoklassicismen inte har abstraherat det sant apokalyptiska blir de sant idealistiska arketypernas efterhängset klaustrofobiska heraldism antidemokratisk.

De sarkastiska ekvilibrismerna kan inte bli koreografiska, enär symbolismen kan bli fullständigt asymmetrisk. Kanske balanserar relativismens geopoliticism de stilistiska fanatismernas metafysik. Arkaismer kan bli harmoniciteter. Det sant profetiskas holism blir det elliptiska. Ickestrategiens tragikomik har troligen agiterats av en eugeni i symbolismen, enär koreografien i det voyeuristiska blir extasens poststrukturalism. Geometrien i hypotoniciteten har inte attraherats av en rytmik, enär de sant pyrotekniska tragikomikerna alternerar de asymmetriska kategorierna i interimismen. Melodier blir amoraliska. De utopiska dekonstruktivismerna i geopoliticismen debuterar inte, emedan intighetens sant labyrintiska kontrapunktik delegerar konstruktivismen i det enigmatiska. De absurdistiska narcissismernas monologism har artikulerats av en analys.

Det atavistiskas helautomatism har avancerat. Det obevekligt absurdistiska förlänar de kameleontiska kvadraturerna de anagogiska jovialismerna. Trots att fanatismens morfologism har agerat plastiken i det materialistiska delegerar det efterhängset symboliskas arkaism fobien. De efterhängset isolationistiska ekosofierna förlänar dialektens fullständigt gigantomaniska plastisk de amoraliska monismernas fullständigt relativistiska autokrati, enär metafysiken i protektionismen förlänar det obevekligt megalomaniska i autokratien det dramaturgiska. Hypokondrier attackerar inte. Kameleontismen avancerar. Naturligtvis förlänar de sant antiheroiska opportunisternas mytologi de fullständigt pedantiska minimalismernas plastisk voyeurismer.

# References

## Literature

- Ahrenberg, Lars. 1990. A Grammar Combining Phrase Structure and Field Structure. In Hans Karlgren (ed.) *Proceedings of COLING-90*, Helsinki, August 1990, Vol. 1: 1-6.
- Bulhak, Andrew C. 1996. *On the Simulation of Postmodernism and Mental Debility using Recursive Transition Networks*. Monash University Department of Computer Science, Technical Report 96/264. Available at:  
<http://www.csse.monash.edu.au/publications/1996/tr-cs96-264.ps.gz>
- Gal, Annie, Guy Lapalme, Patrick St-Dizier and Harold Somers. 1991. *Prolog for Natural Language Processing*. Chichester: John Wiley.
- Pereira, F. C. N. and D.H.D. Warren. 1980. Definite clauses for language analysis. *Artificial Intelligence*, 13:231--278, 1980.

## WWW

(as for Dec 10, 2002)

- ACL www  
<http://www.aclweb.org/>
- Charabia www  
<http://www.charabia.net/generation/index.php?voir=liens&mode=>
- Electrohype www  
<http://www.electrohype.org/electrohype2002/artist.html>
- Horace www  
<http://www.ling.lu.se/persons/Marcusu/misc/horace/index.html>
- Kant www  
<http://macinsearch.com/infomac2/textprocessing/kant-generator-pro-131.html>
- Kurzweil www  
[http://www.kurzweilcyberart.com/poetry/rkcp\\_overview.php3](http://www.kurzweilcyberart.com/poetry/rkcp_overview.php3)
- Lee www  
<http://www.dack.com/web/bullshit.html>
- Postmodernism www  
<http://www.elsewhere.org/cgi-bin/postmodern/>
- Romance writer www  
<http://www.familygames.com/features/humor/romance.html>
- SIGGEN www  
<http://www.dynamicmultimedia.com.au/siggen/>
- Språkbanken www  
<http://spraakdata.gu.se/lb/konk/>
- Toolworx www  
<http://www.burningpress.org/toolbox/>
- Zawinski www  
<http://www.jwz.org/dadadodo/>

## Notes

<sup>1</sup> This confusion between field and constituent structure is good enough for the purposes of this paper. It may be criticized, however; among other things, it invites to circularly defining constituent as "that which fits into a field" and field as "the container of a constituent". Constituents should really be defined independently. See for instance Ahrenberg 1990.