Emergent properties in text generation using tables and grilles

Laura Aylward and Gordon Rugg

Abstract

Gordon Rugg's 2004 Cryptologia article showed that using text generation tables combined with modified Cardan grilles can produce text with complexity and structural properties similar to those of the Voynich manuscript. A suggested topic for further work was investigation of the properties arising from tables more highly structured than those used in that article. This article describes our recent work in this area. It:

- clarifies some common misunderstandings about the table and grille method
- describes findings from our work on highly structured tables, and
- tests some hypotheses arising from this work

We conclude that several unusual properties which occur in the Voynich manuscript arise naturally as emergent properties from using highly structured tables in conjunction with modified Cardan grilles. These include the following:

- a high proportion of "blank" words, leading to variable line lengths
- different frequencies of the same syllable in different parts of the table
- sections consisting mainly of long words alternating with sections consisting mainly of short words, and
- words which are common in text produced using one grille being absent from text produced using a different grille on the same table

Some side-effects of this method would, unless corrected, be easily visible in a meaningless hoax document. Correcting these side-effects will in turn result in different regularities in the output text. One side-effect, mentioned briefly in the Cryptologia article, is that the movement of grilles across a table has to include a significant proportion of random movement, to break up regularities in syllable patterns. A related issue is that the first word on each manuscript page has to be generated in a different way from "normal" text, to obscure regularities which would otherwise be apparent.

We are examining the manuscript to see whether these features are present, and will report our findings in a future document.

Introduction: some common misunderstandings

The Cryptologia article describes how text generation tables can be used in conjunction with modified Cardan grilles to produce text with similar properties to Voynichese. This demonstrated that it was possible to produce text with similar complexity to Voynichese as a meaningless hoax, implying that the Voynich manuscript could have been hoaxed. That is a separate question from whether the manuscript was hoaxed. It is also a different question from whether the manuscript was produced using this method, either as a hoax or as a ciphertext.

An argument has been made that if the manuscript were produced using tables and grilles, then it would be possible to reconstruct one or more pages of the manuscript through reconstructing the original tables and grilles. This argument is based on a

misunderstanding of how tables and grilles are used in the production of meaningless hoaxes. This section provides a brief summary of the method to clarify this point, and to provide background for some of the points raised below.

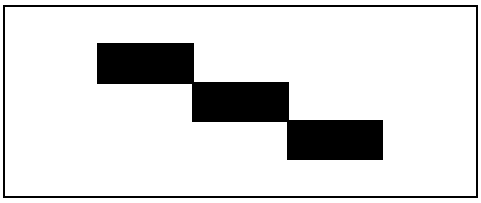
The table and grille method produces Voynichese-like text by using a large table of meaningless syllables (about 40 rows by 39 columns) which are combined by using a modified Cardan grille. The grille is moved semi-randomly across the table. For simplicity, we have used Stolfi's breakdown of Voynichese words into three components, namely prefix, midfix and suffix, although tables can be used with other breakdowns if desired. (We use the term "midfix" rather than "infix", with "midfix" referring specifically to the middle slot in Stolfi's model. There are other models of Voynichese which go beyond the scope of this document, and for which "infix" would be used differently from "midfix".)

The table fragment below illustrates how the method is used. For clarity, we have used different colours to show how each sequence of three columns fits together, with a sequence of prefix, midfix and suffix columns. For instance, the first column contains prefixes such as "qo" in red, the next column contains midfixes such as "chek" in red, and the third column contains suffixes in red. This pattern is repeated in the next set of three columns, shown in blue italics, and again in the third set of three columns, in green bold characters. (The tables we actually use are monochrome –we have coloured the cells below to help illustrate how the method works.) Some cells are deliberately blank.

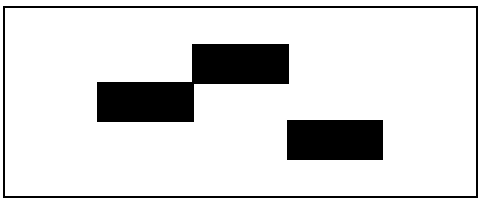
qo	chek		*ol*	*she*	*dar*	**sol**		**dy**
	she	dy	*qo*	*kee*	*dy*		**she**	**y**
y	k	y		*tee*	*y*	**qo**	**ke**	**dan**
ol		dor	*y*			**dy**	**k**	
qo	t	dy	*or*	*che*	*dy*	**o**	**che**	**dor**

A grille is a piece of card, with slots cut in it. Two examples are shown below.

Grille 1 has three slots in a diagonally descending sequence.



Grille 2 has three slots in a "rise and fall" sequence.



If grille 1 is placed over the red cells of the table fragment, with its prefix slot on the top row, it will reveal the word “qoshey”. If it is now moved to the blue cells it will reveal the word “olkeey”, and if moved straight across to the green cells it will reveal “solshedan”, making the “sentence” “qoshey olkeey solshedan”. When moved to the second row, it produces the words “kdor qotee ke”.

If we repeat this with grille 2, the first line we produce is “cheky qoshey y”, and the second is “yshedor kee qoshe”.

At first sight these look like quite different sets of output, but if we put the first line from grille 1 next to the second line from grille 2 and tabulate them, then a regularity becomes apparent.

grille1	qo	she	y	*ol*	*kee*	*y*	**sol**	**she**	**dan**
grille2	y	she	dor		*kee*		**qo**	**she**	

Although the grilles have combined the syllables into different words, the midfixes in these two lines are the same in consecutive words (and, likewise, the prefixes and the suffixes derived from the same row of the table will occur in the same sequence unless something is done to prevent this). One simple way of breaking the regularities is to introduce some randomness into the grille’s movement – for instance, by skipping from the red columns directly to the green, without generating a word from the blue columns, or by moving down two rows part-way across the table, or by moving to a completely different part of the table. Other ways include introducing an arbitrary division within the word generated – for instance, breaking “qochedy” into “qoche” and “dy” – and introducing extra words such as “daiin”, or concatenating two successive words.

This means that the same table and grille, used in this way, will produce different output each time they are used.

Introducing a random movement two or three times per line would mean that the longest regular sequences would consist of only a few items (for instance, four words containing the same sequence of midfixes). Given the highly repetitive nature of Voynichese, it would be very difficult to distinguish between a “genuine” repetition produced by two different grilles moving across the same section of table, and a “spurious” repetition, where the same sequence of (say) midfixes occurred by chance. A possible exception would be when a very rare character is used, and are currently examining the manuscript for such cases.

Levels of structure

Tables with different degrees of structure produced text with different properties. In brief, the lower the degree of structure, the more consistent the word length. With highly structured tables, word lengths were very variable, and there was a high proportion of “blank” words, where the grille revealed only empty cells in the table. Some examples of different levels of structure are given below; blank words are identified by two consecutive full stops [..]. The rest of this document focuses on highly structured tables unless otherwise specified.

Table 3.1: Example output texts from five tables with a range of structure levels.

Table Structure: Low
olky.qochedy.keain.qokey.qoshdy.qokey.ykdy.okal.keol.shey.
dlshey.qotd.olshey.lchey.qotdy.y.sheeaiin.salkol.dkeaiin.oshey.
okear.qotar.ky.qoshey.shedaiin.qokm.ochy.tedy.qokeedy.teedy.
qoshecthdy.dy.tey.qoaiin.okdy.y.y.lcheaiin.chey.qotedy.
qotchey.keeaiin.qodtdy.olcheeor.cheain.olshckhy.qody.okeeedy.lkdy.kea
i
from: stolfiLowOutput6.txt
Table Structure: Medium-Low
ochel.oqoshealdy.qok.yshey.qochedy.qoshey.chcthdy.ly.ycheol.chedy.
olchear.qofchey.chekdy.oshekdy.qoqocthey.qokeeol.key.kedy.kdy.teey.
qoor.qoshedy.ldtdy.dolsheey.kedy.qoshy.qody.kaiin.qoche.oky.
lky.okey.teol.chey.qokeor.qotalol.qoldy.oldy.qopdy.qody.
olkol.lsheal.qosher.shedy.kol.kdy.qokol.dy.olchedy.keol.
from: stolfiMLowOutput6.txt
Table Structure: Medium
qolkeear.kar.oky.qokeey.kdy.qolkal.otear.shey.cheaiin.tdy.
okdy.cheaiin.qokeedy.kdy.qoshear.olkol.oshey.ke.lshedy.keeal.
keedy.oky.olchdy.keedy.okeol.qoshey.che.tdy.qotdy.kdy.
sheol.kedy.ky.qochedy.lkeey.otain.qotdy.okee.sheaiin.keaiin.
qokdy.chedy.keeaiin.ochey.olcheol.stedy.qolks.qotee.keol.lshe.
from: stolfiMedOutput6.txt
Table Structure: Medium-High
aiin.tdy.qokeedy.y.qoaiin.otey.kdy.ky.lcheey.qochedy.
t.kedy.qochey.qody.oshe.qokdy.ok.qochey.qotaiin.keedy.
keedy.chy.qoche.o.dy.kol.chey.qotdy.kedy..
ky.qosheaiin.okal.qody.dk.ty.qoshear.keedy.oche..
qo.oshear.oldy.kal.rkey.qoaiin.chckhdy.y..y.
from: stolfiMHighOutput6.txt
Table Structure: High
dy.qoche.ok.kedy.ly.qo.shedy..ky.qochedy.
qoche.ody..ky.qoshedy..kee.shdy.qoche.ok.
o.kee.shedy.qoaiin.olkol.kedy.y.qoche.ody..
shedy.lteey.qo.tdy..ky.qochedy.oaiin.kee.shedy.
ky.qosheedy..keeal.dy.qoche.ok.shedy.ly.qot.
from: stolfiHighOutput6.txt

Findings from use of highly structured tables

There are various ways of filling in the cells in the table. At one extreme, each set of columns can be filled with the relevant type of syllable (e.g. prefixes) in a random or quasi-random manner. At the other extreme, they can be filled in a highly structured manner (for instance, a “qo” in every fifth prefix cell). The findings reported in the Cryptologia article derived from randomly populated tables. There were obvious limitations to this, so we have followed up that work via software written by Laura Aylward which allows us to manipulate the degree of structure. The most interesting findings came from highly structured tables, which we report below.

A significant point is that these properties emerge spontaneously from highly structured tables. The previous Cryptologia article showed that these properties could be produced by someone who used tables and grilles in a particular way, either by design or by accident. Our more recent work shows that these properties are not just explicable via tables and grilles, but are to be expected: for instance, using highly structured tables and grilles can be expected to produce two different “dialects”.

Different text from same grille with different movements

An implication of the need for random movement in grilles is that using the same table with the same grille on different occasions will lead to the production of different output text. We include some examples of this, where the same table and grille have been used with different movements, and produced different outputs, even with highly structured tables.

This makes attempts to reconstruct underlying tables difficult or impossible, although it may be possible to reconstruct fragments of tables.

The table below shows output from using the same grille with different patterns of movement on a table with a medium degree of structure. Pattern 1 involves moving the grille horizontally across the table; pattern 2 involves moving the grille diagonally at intervals across the table, pattern 3 involves moving the grille across the table and then randomly up or down the table after each horizontal movement. We include two examples of pattern 3. For simplicity, pattern 3 uses a regular horizontal movement of the grille across the table, though a more realistic model for a hoax would involve a degree of both random horizontal and random vertical movement.

All four text samples were produced from the same starting point on the same table with the same grille design.

1	1	**stolfiMedOutput1.txt** keey.shey.chey.oky.qochey.qokey.ochey.qokchy.ochey.ochy. qotol.qok.shey.teaiin.shedy.rtal.k.qokeaiin.qok.che. tdar.t.tdy.chear.qokar.qoky.she.okeey.chedy.kdy. kee.ky.qoke.chey.olshe.tedy.keeaiin.ocheal.kaiin.qoshey.
	2	**stolfiMedOutput5.txt** keey.qok.tdy.chey.qokdy.qokee.oltdy.sheaiin.qochey.chdy. qotol.t.qoke.chedy.oty.k.k.chey.qokdy.qok. tdar.ky.qotdy.qotdy.t.qoches.tey.shedy.okeaiin.kdy. kee.tdy.shedy.kaiin.qoky.ky.okdy.kee.lkaiin.qote.
	3	**stolfiMedOutput9.txt** kee.ky.shedy.chey.qokdy.kdy.oltdy.lshedy.chedy.kdy. qotol.ky.shedy.chedy.olshe.kdy.keeaiin.qokeaiin.qotdy.keedy. qosheal.okaiin.shedy.chey.olshe.qokee.keeaiin.ocheal.dshedy.keedy. chedy.lchedy.qotdy.chedy.t.tedy.qokedy.qochedy.kaiin.qoshey.
		stolfiMedOutput10.txt qotol.tdy.shey.qotdy.qokar.tedy.oltdy.ocheal.qok.kdy. okeedy.okaiin.tdy.qotdy.oty.tedy.she.qochedy.chedy.chedy. kedy.ky.qoke.chey.qokar.qokee.chey.okeey.qotdy.kdy. qochedy.lchedy.shedy.chey.oty.k.keeaiin.chey.kaiin.qok.

The two samples produced using random movements (stolfiMedOutput9.txt and stolfiMedOutput10.txt) are markedly different from each other, even though they were produced using the same table and the same grille. This illustrates the difficulties in trying to recreate tables from output text, and also shows that attempts to recreate the Voynich manuscript word for word would probably not succeed, even if original tables and grilles were found.

Initial words

Another manifestation of the “regular sequence” effect is that if the grille is moved regularly across the table, starting at the top left of the table, then the first word in each batch of output text will contain a very limited set of prefixes, midfixes and suffices, reflecting where the grille slots are positioned. If the grilles consist of four possible positions for the prefix slot (row 1, 2, 3 or 4), for instance, then the first word of each set of output will always begin with one of the same four prefixes.

If this method is being used to produce a meaningless hoax, then it is necessary to obscure this effect in some way. Various methods are possible.

The implication is that the first word of each page of output produced using this method will be different in some way from words elsewhere in the text. This is consistent with Stolfi's finding of unique initial words for the pages he studied.

An example of this is shown above, where stolfiMedOutput1.txt and stolfiMedOutput5.txt both begin with the syllable "kee", as does one of the outputs using random movements.

Blank words

A surprisingly high proportion of the text generated with highly structured tables (about 10%) consisted of "blank" words, i.e. words where the prefix, midfix and suffix slots revealed by the grille were all empty cells.

The implication is that it is not possible to make an unambiguous mapping between output text and the table used, even if there is no attempt to conceal the sequence in which the grille is moved across the table. For instance, if the second word revealed by a grille is a blank, then there will be no corresponding word visible in the output, and the second word on the written line of manuscript will have been produced by the third section of the table. (For example, the blue section of table above might produce a blank word, so the output text would apparently jump from a "red" word to a "green" word.)

An example of text produced with a highly structured table is shown below. Note that every line contains at least one blank word, indicated by two consecutive dots, and that the second line contains two blank words. Line three contains a blank word at the end of the line.

Table Structure: High

```
dy.qoche.ok.kedy.ly.qo.shedy..ky.qochedy.
qoche.ody..ky.qoshedy..kee.shdy.qoche.ok.
o.kee.shedy.qoaiin.olkol.kedy.y.qoche.ody..
shedy.lteey.qo.tdy..ky.qochedy.oaiin.kee.shedy.
ky.qosheedy..keeal.dy.qoche.ok.shedy.ly.qot.
```

from: stolfiHighOutput6.txt

Variable word and line lengths

Word lengths turned out to be highly variable in a way which appeared non-random (and probably was indeed non-random). This meant that a line of output might consist entirely of short words, including one or more blank words.

The implication is that if this method was used to produce text for a manuscript, then the table would sometimes not produce enough text to reach the end of a line of the manuscript; this in turn implies that some method would need to be used to provide text to complete the line. Various methods are possible; all of them would probably

lead to the introduction of “noise” in terms of character distributions across the output text, and would complicate any attempt to reconstruct underlying tables. For instance, a “short” line could end with an “m” in the normal manner (if “m” is in the last column of the table), and the line could then be extended to full length by adding extra text; this would result in an “m” appearing in the middle of the line. The extra text could come from a random position on the table, or from starting again on a new line of the table, which would further obscure any regularities in character distributions in the table. A hypothetical example is shown below, with the “extra” text in italics. Note the blank word, indicated by “..” between “daiin” and “or”, and the position of the “skam”.

oky.oly.daiin..or.sor.dy.or.dain.skam.*qoky.saiin.sor*-

The table below shows text produced using different levels of table structure. The tables with low structure in this example tend to produce words which are of more consistent length and which are usually longer than those from highly structured tables.

Table Structure: Low

olky.qochedy.keain.qokey.qoshdy.qokey.ykdy.okal.keol.shey.
dlshey.qotd.olshey.lchey.qotdy.y.sheeaiin.salkol.dkeaiin.oshey.
okear.qotar.ky.qoshey.shedaiin.qokm.ochy.tedy.qokeedy.teedy.
qoshecthdy.dy.tey.qoaiin.okdy.y.y.lcheaiin.chey.qotedy.
qotchey.keeaiin.qodtdy.olcheeor.cheain.olshckhy.qody.okeeedy.lkdy.kea
i

from: stolfiLowOutput6.txt

Table Structure: Medium-Low

ochel.oqoshealdy.qok.yshey.qochedy.qoshey.chcthdy.ly.ycheol.chedy.
olchear.qofchey.chekdy.oshekdy.qoqocthey.qokeeol.key.kedy.kdy.teey.
qoor.qoshedy.ldtdy.dolsheey.kedy.qoshy.qody.kaiin.qoche.oky.
lky.okey.teol.chey.qokeor.qotalol.qoldy.oldy.qopdy.qody.
olkol.lsheal.qosher.shedy.kol.kdy.qokol.dy.olchedy.keol.

from: stolfiMLowOutput6.txt

Table Structure: Medium

qolkeear.kar.oky.qokeey.kdy.qolkal.otear.shey.cheaiin.tdy.
okdy.cheaiin.qokeedy.kdy.qoshear.olkol.oshey.ke.lshedy.keeal.
keedy.oky.olchdy.keedy.okeol.qoshey.che.tdy.qotdy.kdy.
sheol.kedy.ky.qochedy.lkeey.otain.qotdy.okee.sheaiin.keaiin.
qokdy.chedy.keeaiin.ochey.olcheol.stedy.qolks.qotee.keol.lshe.

from: stolfiMedOutput6.txt

Table Structure: Medium-High

aiin.tdy.qokeedy.y.qoaiin.otey.kdy.ky.lcheey.qochedy.

t.kedy.qochey.qody.oshe.qokdy.ok.qochey.qotaiin.keedy.

keedy.chy.qoche.o.dy.kol.chey.qotdy.kedy..

ky.qosheaiin.okal.qody.dk.ty.qoshear.keedy.oche..

qo.oshear.oldy.kal.rkey.qoaiin.chckhdy.y..y.

from: stolfiMHighOutput6.txt

Table Structure: High

dy.qoche.ok.kedy.ly.qo.shedy..ky.qochedy.

qoche.ody..ky.qoshedy..kee.shdy.qoche.ok.

o.kee.shedy.qoaiin.olkol.kedy.y.qoche.ody..

shedy.lteey.qo.tdy..ky.qochedy.oaiin.kee.shedy.

ky.qosheedy..keeal.dy.qoche.ok.shedy.ly.qot.

from: stolfiHighOutput6.txt

Different frequencies of the same syllable in different parts of the table, as an effect of how the table is filled in

Highly structured tables involve allocating syllables to table cells in a highly systematic manner – for instance, putting a "qo" in every fifth prefix cell. This can lead to complications after a significant proportion of the table has been filled, particularly if the process works systematically from (say) the shortest syllables to the longest ones and/or the rarest to the commonest. The result can be different frequencies of the same syllable in different parts of the table (e.g. the left versus the right side, or the top versus the bottom).

For instance, if the table is filled starting with short common syllables, and ending with rare long syllables, then by the time the rare long syllables are reached, many of the slots where they would otherwise have been placed will already be filled. For example, if the syllable "oldy" was to be placed in every seventh suffix cell, then cell 7 would probably be free, but cell 35 might already be occupied by a syllable which filled every fifth suffix cell, and cell 42 might be occupied by one which filled every sixth suffix cell. This would mean that "oldy" would be rarer towards the end of the table than the beginning. If the table is filled in column by column, top to bottom, left to right, then "oldy" would be more common on the left of the table, and would be more frequent towards the start of lines produced using that table (other things being equal, such as grille movements).

There is also a tendency for rare characters to appear in the same row or column, as a result of regularities in how table is filled in; this would be exaggerated if randomness in grille movement was added by using vertical movement of the grille up and down the table, and reduced or obscured if randomness was introduced by using horizontal movement of the grille across the table, either forwards or backwards.

Alternating sections of long and short highly repetitive words

Partly as a result of the effect described above, highly structured tables can produce text which contains alternating sections of long, highly repetitive words and short, highly repetitive words.

The example below shows this effect in a single set of output from a single table and a single grille.

ody.o.o.ochedy.ok.o.oky.ok.oke.ocheaiin.
dy.dy.oldy.keedy.dy.dy.chedy.dy.tedy.dy.
kee.qoy.o.qo.dy.dy.dy.o.she.chedy.
y.qochey.keey.y.y.qoy.y.ry.shey.key.
qoal.qoaiin.qoshear.qosheol.qocheaiin.qoshey.qo.qosheaiin.qoshe.qo.
kdy.keedy.keedy.dy.kedy.dy.dy.olchedy.dy.dy.
k.kdy.kdy.olkol.kar.ky.k.k.olkol.k.
keedy.y.che.d.dy.ar..kee.qolkedy.k.
chey.lshey.qoy.ty.shey.chey.y.shey.qotey.tey.
qody.qokeedy.qochedy.qokeedy.qody.qokeedy.qotedy.qokeedy.qochedy.qokeedy.

Different "dialects" from different tables

As expected, different tables produced different outputs. Interestingly, though, different "dialects" were produced by tables with different amounts of structure, even when the tables used the same syllables in the same proportions. For instance, in one batch of outputs, the word *'qoky'* appeared in text generated from the low structure table, yet in no other text using the same syllables in the same proportions, but using different degrees of structure.

This has clear similarities with the "dialects" of Voynichese.

Discussion

If highly structured tables are used in conjunction with modified Cardan grilles to produce a meaningless hoax, then various features arise naturally in the text which is produced. Some of these features are easily detectable unless deliberately obscured, and would be obscured by a reasonably competent hoaxer.

Text produced in this way shows the following features:

- sequences of repetitive short words alternating with sequences of repetitive long words (this is reduced if the tables are less structured)
- words common in one set of output text being completely absent from other sets of output text (if the degree of structure in the tables is changed)
- rare characters tending to occur in a limited number of places on the page
- rare characters tending to occur in many unique words, rather than in repeated instances of the same rare word (since each grille will have a unique pattern, and will combine a rare character with different syllables each time it is encountered)

The features which would presumably be deliberately obscured in the production of a meaningless hoax are

(a) the regular sequences of syllables which would arise if the grille was moved systematically across the table and
(b) the tendency for first words on a page to consist of a very limited syllable set.

In addition, systematic movement of the grille would lead to very rare characters always occurring in predictable positions in a line of text, so this would presumably also be deliberately obscured.

Conclusion

Previous work with randomly constructed tables and Cardan grilles showed that text with most of the striking features of Voynichese, including apparent complexity, could be produced fairly easily. Using highly structured tables with modified Cardan grilles in the way described in this article has shown that these features, plus some others not described in the previous article, arise naturally as emergent properties from the structure of the tables.

Part II of this document is in preparation, and will examine features of the manuscript to see whether there is evidence for or against the hypothesis that the manuscript actually was produced using tables and grilles, as opposed to whether it could have been.