Title
Structural Complexity in Fairy Tales

Author
Lakoff, George

Publication Date
1972

Peer reviewed

The Study of Man, Vol. 1 (1972), 128-150.

Structural Complexity in Fairy Tales

George P. Lakoff*

PREFACE

This is an inquiry into the structure of discourse through the investigation of an interesting subset of discourses -- Russian Fairy Tales. I assume that the study of discourse is ultimately to be part of a formalized theory of linguistic structure. I shall endeavor to show that any adequate theory of the structure of fairy tales must share many formal properties with the transformational linguistic theory that has so far been proposed by Noam Chomsky and his co-workers. Such a result would be of psychological interest, for it would indicate that an adequate model for the hearer and speaker of the sentences of a language could also be used to describe, at least in part, the human ability to produce and understand discourses. This might ultimately show that the human mind is an extremely efficient device which uses essentially the same mechanism for constructing and understanding complicated discourses as it uses for constructing and understanding individual sentences.

---

*This paper was written in 1964 while my remote ancestor, George P. Lakoff, was a graduate student at Indiana University. I trust that that eponymous worthy would take it as no disservice that I am unable to find myself in total agreement now with what he wrote then; though there is a certain melancholy pleasure, I own, in re-reading even those passages where we most differ. ---George Lakoff, University of California, Berkeley.

My short-term goal is to alert students of highly stylized discourses -- folklorists, literary scholars, and some psychologists and philosophers -- to the probability, if not the fact, that all coherent human discourse is intricately structured and that some of the formal tools capable of representing such structures have already been made available in the study of the foundations of linguistic theory. Moreover, I offer the suggestion that the study of the structure of stylized discourses will ultimately be most revealing not as an ad hoc classifying procedure (such as a compilation of tale types or a neo-Aristotelian literary taxonomy), but as an attempt to discover in what formal properties stylized discourses coincide with all human discourse, and in what particular formal features they diverge.

I will attempt to demonstrate, as Chomsky has done for English sentences, that Finite State and Phrase Structure models are inadequate for the description of fairy tale structure and that only a transformational model has sufficient formal power to describe accurately the complexities of the structure of fairy tales.

I

A discourse is essentially a string of English sentences. But not every string of sentences forms a discourse. Take Newmark and Bloomfield's example of a discourse and a non-discourse[1]:

(1) Yesterday I was in this bar. The funniest thing happened. A lady walked in and asked for the manager in a loud voice. Well, he came over and asked what the trouble was. All she said was, "You Satan," and then she hit him over the head with her wet umbrella.

(2) All she said was, "You Satan," and then she hit him over the head with her wet umbrella. Well, he came over and asked what the trouble was. A lady walked in and asked for the manager in a loud voice. The funniest thing happened. Yesterday I was in this bar.

They comment:

> Passage (1), a discourse, makes good linguistic sense, though perhaps not much intellectual sense. The reader or listener follows it from one part to the next, and even though he has no idea of the significance of the events described, he is willing to accept that the writer or speaker means to describe these events as a whole. In contrast, passage (2), a mere collection of sentences and no discourse, makes no linguistic sense; the reader or listener can make nothing out of the strange sequence. It could not stand on its own as a discourse to be interpreted in its own right, as passage (1) could.

In order to make "good linguistic sense," a string of sentences must have at least two properties: it must be connected and it must be structured. In order to be connected, a string of sentences must conform to certain grammatical restrictions, which are as yet very imperfectly known. For example, pronouns must refer properly to their antecedents in previous sentences. The definite article is used when reference is made to someone or something in a previous sentence. By reversing the order of (1) to form (2), we have upset the agreement restrictions on pronouns and articles. We will not concern ourselves with connectedness here. Our main problem will be to study the internal organization of discourses -- or how they are structured.

To get at the differences between connectedness and structure, let us consider two strings of sentences which are connected but not structured.

(3) Little Johnny wanted a bicycle. Bicycles were invented by Abner Doubleday in 1776. In that year the Charles River overflowed, drowning two flea circus entertainers in Canton, Ohio. Ohio's manure industry provides thirty-eight percent of the state's gross revenue. Gross earnings of professional tennis players are rising.

(4) Little Johnny wanted a bicycle. So he started delivering newspapers.

The speaker in (3) rambles on from sentence to sentence, changing topics as he goes along. Although each sentence connects with the previous one, nothing he says "fits together." In (4) we are left hanging in mid-air. We expect more. We feel that (4) is the beginning of a story and we are waiting to hear the rest of it. Such a feeling reflects our intuition that stories are structured, that they have certain constituent elements, and that in (4) only some of the initial elements are presented. It is like the feeling that we have when we hear, "A big grey..." and we expect a concrete noun to follow, because we know implicitly how English sentences may or may not be structured. It may be the case that just as we know the grammar of English, we may in some sense know the "grammar" of story construction.

II

Modern linguistic theory is founded on the concept of the linguistic level. Each level -- phrase structure, morphemics, phonemics -- is essentially a device for describing a different kind of structure; each is a different way of representing utterances. The importance of V. Propp's Morphology of the Folktale is that in it Propp has isolated a significant level of analysis for fairy tales.

On the surface, fairy tales, like all discourses, are strings of sentences. However, sentences are not appropriate units for representing the internal structure of fairy tales and other discourses. After all, entire tales or large parts of them can be told in single sentences. For instance,

> Ivan, having received word of his father's death at the hands of a dragon, went forth seeking revenge and encountered an old man, who, after asking him three riddles which he successfully answered, gave to him three magical gifts, the first of which, a winged horse, took him to the dragon's castle, the second of which, a universal master key, opened the castle gate, and the third of which, a magic sword, enabled him to slay the dragon and thus avenge his father's death.

Obviously, sentences will not do as structural units. Some other units are necessary, abstract ones, ones which cannot be explicitly found in the data and must be postulated.

The usual unit postulated, if any is postulated at all, is the motif. Propp rightly rejects this as a plausible structural unit.

> If a motif exists as something logically whole, then each sentence of a folktale represents a motif. (A father has three sons: a motif; a stepdaughter leaves home: a motif; Ivan fights with a dragon: a motif; and so on.) All this would be perfect if motifs were really indissoluble; an index of motifs would then be made possible. But let us examine the motif "a dragon kidnaps the king's daughter" (this example is not Veselovskij's). This motif may be composed into four elements, each of which, in its own right, is capable of variation. The dragon may be replaced by Koscej, a whirlwind, a devil, a falcon, or a sorcerer. Abduction can be replaced by vampirism or any other means by which disappearance is effected in folktales. The daughter may be replaced by a sister, a bride, a wife, or a mother. The king can be replaced by a king's son, a peasant, or a priest. In this way, contrary to Veselovskij, we must affirm that a motif is not monomial nor decomposable. The last decomposable limit, as such, does not represent a logical whole.[2]

Propp's reasoning as to structural units proceeds like this: consider the following events:

1. A king gives an eagle to a hero. The eagle carries the hero (the recipient) away to another kingdom.
2. An old man gives Sucenko a horse. The horse carries Sucenko away to another kingdom.
3. A sorcerer gives Ivan a little boat. The boat takes him to another kingdom.
4. The princess gives Ivan a ring. Young men appearing from out of the ring carry him away into another kingdom.[3]

In these sequences only objects and dramatis personae are variables. The actions remain the same. The constant elements, then, must be actions, not the people who perform them nor the objects or persons on whom they are performed. "Definition of a function is most often given in the form of a noun expressing an action (interdiction, interrogation, flight, etc.)."[4] But, Propp goes on, the same actions are not always examples of the same structural units.

> An action cannot be defined apart from its place in the process of narration. The meaning which a given function has in the process of action must be considered. For example, Ivan's marriage to the king's daughter is something entirely different from the marriage of a father to a widow with two daughters. A second example: if, in one instance, the hero receives money from his father in the form of one hundred rubles and subsequently buys a magic horse with this money, and, in a second case, the hero is given a sum of money as a reward for an accomplished act of bravery (at which point the tale ends) we must admit to the existence here of two morphologically different elements -- in spite of the apparently identical action (the transference of money) in both cases. Identical acts can have entirely different meanings and vice versa. Function must be taken as an act of the dramatis personae, which is defined from the point of view of its significance for the course of action of a tale as a whole.[5]

A function is then a set of a finite number of actions that can occur at a given point in a tale. A given action may belong to more than one function.

Having defined his structural units, Propp goes on to propose the following theory of fairy tale structure:

1. Functions serve as stable, constant elements in folktales, independent of who performs them, and how they are fulfilled by the dramatis personae. They constitute the components of a folktale.
2. The number of functions known in the fairy tale is limited.
3. The sequence of functions is always identical.

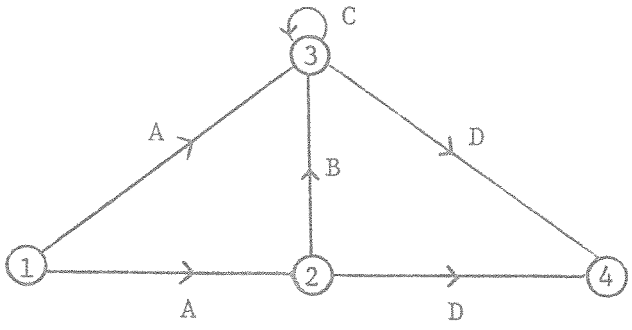
> As for groupings, one ought first of all to realize that far from all folktales give evidence of all functions. Yet this does not change the law of sequence for one moment. The absence of several functions does not change the order of those remaining.

Essentially, Propp claims that the structure of a folktale can be adequately described by a sequence of functions, some of which may be absent.

Let us ask what kind of formal theory will give us an adequate representation of the set of such sequences; or equivalently, of the set of structural descriptions that Propp proposes for fairy tales. An equivalent question is to ask what kind of device we can construct that will generate all possible sequences of functions.[7]

One such device is a finite state automaton: in essence, a machine that can be in any one of a finite number of different internal states. Suppose the machine switches from one state to another by producing a certain symbol, say the symbol for a function. One state is the initial state, another is the final state. Let the machine begin in the initial state, go through a number of other states producing a function with each change, and end in the final state. What kind of machine of this sort will produce the sequences of functions that Propp claims will describe all possible fairy tales?

Propp gives thirty-one functions for Russian fairy tales. We can limit ourselves to four in our description with no loss of generality. Let the functions be A, B, C, and D. Suppose A and D must be present, B and C being optional. Then, a device for generating the appropriate sequences of functions would be represented by the following state diagram:



A finite state device of this sort can generate all the possible sequences of functions in Propp's statement of his theory. (In the above diagram, numbers indicate states, arrows indicate possible state transitions, and the letter next to each arrow indicates the function produced when that state transition occurs.)

The structure of folktales would be fairly simple to describe if this theory were adequate. Unfortunately it isn't. In fact, given facts that Propp himself points out about the structure of tales, we can show that the set of structural descriptions cannot be generated by a finite state automaton.

Our argument follows the one that Chomsky puts forth in Syntactic Structures (21-22) to show that English sentences cannot be generated by finite state devices. We will consider three sets of strings of elements which cannot be generated by finite state automata. Then we will show that indefinitely many structural descriptions of fairy tales must contain strings of the sort found in each of these sets.

Chomsky cites the following sets of strings composed of elements a and b:

(5) (i) ab, aabb, aaabbb, ..., and in general, all sentences consisting of n occurrences of a followed by n occurrences of b and only these;

(ii) aa, bb, abba, baab, aaaa, bbbb, aabbaa, abbbba, ..., and in general, all sentences consisting of a string X followed by the 'mirror image' of X (i.e., X in reverse), and only these;

(iii) aa, bb, abab, baba, aaaa, bbbb, aabaab, abbabb, ..., and in general, all sentences consisting of a string X of a's and b's followed by the identical string X, and only these.[8]

It has been shown that these sets are not capable of being generated by a finite state device.[9] In general, a finite state device can handle only strings which are put together by placing one element after another consecutively. Such a device cannot handle strings which are formed by embedding strings within other strings an indefinite number of times.

Fairy tales have the property that one tale (sequence of functions) may be embedded within another, which is embedded within another, and so on. We are all familiar with tales in which the hero sets out to avenge

one villainy and, on his way, has adventures which are tales in themselves, and then ultimately accomplishes what he originally set out to do. Such a tale, one with others embedded within it, is essentially of the form (5) (ii) and a description of it cannot be generated by a finite state device. Moreover, tales in which whole sequences of actions are exactly repeated two or three times have structures of the form (5) (iii) and cannot be handled by a finite state device.

Thus, the model which Propp explicitly proposed is not adequate to describe tales. Propp, however, was not chained by his own theory, and the most interesting part of his work is his description of plot structures which cannot be accounted for by the model described above and which assume a more powerful theoretical framework for their description.

III

> ...a large number of functions are arranged in pairs (prohibition-violation, investigation-distribution, struggle-victory, persecution-deliverance, etc.). Other functions may be arranged according to groups (villainy, dispatch, decision for counteraction and departure from home, for example, constitute a start of the plot. Elements EDG, as well, form something of a whole).
>
> -Propp, 58

> Morphologically, a folktale may be termed any development out of villainy (A), or a lack (a) through intermediary functions to marriage (Rs), or to other functions used in the capacity of the dénouement. Terminal functions are, at times, a reward (F), a gain or the general liquidation of misfortune (K), a rescue from pursuit (W*), etc. This type of development is termed by us as a move. Each new villainy, each new lack, creates a new move. One folktale may have several moves....
>
> -Propp, 83

Propp's conception of fairy tale structure as it ultimately evolves in his book is that a tale with a complicated plot can be decomposed into a number of tales with simple plots. These simple plots, or "moves," are all of the same form: a string of elements which are to be arranged in pairs or groups. The breaking down of strings of elements into groups is equivalent to the descriptive device in linguistics

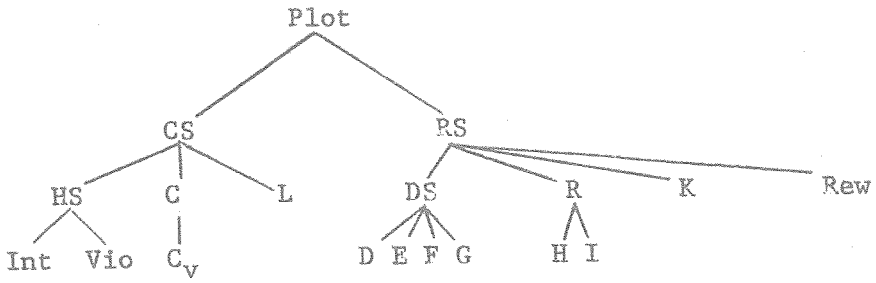
known as constituent analysis, or parsing. As Chomsky has shown[10], the model presupposing such a descriptive device is essentially more powerful than the finite state model which is equivalent to Propp's descriptive statement. A device capable of constituent analysis is called a Phrase Structure Grammar, and can be described by a list of rules of the form: A ⟶ B+ .... +C; this is to be read "A is rewritten B+...+C." In other words, one symbol is to be rewritten as a sequence of symbols.

Consider the following example of the operation of such a grammar:

(6)
1. Plot → CS+RS
2. CS → HS+C+L
3. HS → Int+Vio
4. RS → DS+R+K+Rew
5. DS → D+E+F+G
6. C → C_V
7. R → H+I
8. Int → Ivan is warned not to leave his sister alone in the house.
9. Vio → Ivan violates the warning.
10. C_V → A dragon kidnaps his sister.
11. L → Ivan discovers the misdeed and leaves in pursuit.
12. D → Ivan encounters an old man who asks him a riddle.
13. E → Ivan answers correctly.
14. F → The old man gives Ivan a horse and a sword.
15. G → The horse takes him to the dragon's kingdom.
16. H → Ivan fights the dragon.
17. I → Ivan kills the dragon with the sword.
18. K → Thus, Ivan rescues his sister.
19. Rew → Ivan is awarded the 4-H Club heroism medal.

Terms:
CS: Complicating Sequence
RS: Resolving Sequence
HS: Helplessness Sequence (Here the hero is reduced to helplessness prior to a villainy.)
DS: Donor Sequence
Int: Interdiction
Vio: Violation
R: Resolution
C: Complication (Propp's A)
C_V: Villainy
D: (Propp's term) Donor interrogates, tests, or attacks hero.
E: " " Hero reacts appropriately.
F: " " Hero receives magical agent.
G: " " Hero uses magical agent.
H: " " Hero fights villain.
I: " " Hero defeats villain.
K: " " Misfortune is liquidated.
Rew: Reward

The sentences on the right of the arrow from 8 through 19 constitute the tale. The tree diagram below (7) indicates the tale's constituent structure. Each node represents a constituent. Notice that the tale has a hierarchical structure.



In (6) we rewrote the functions (we will henceforth call them Action Categories, or A-categories) as sentences for the sake of convenience. As Propp says, they should be written as actions in some abstract form, with blanks for filling in the subject and object. Ideally, the abstract form for representing such actions would be taken from an adequate theory of semantic structure. Since none exists, we can only employ an ad hoc method for representing the actions and their subjects and objects. The following device is chosen for no reason other than convenience in stating rules.

(8) Associate with every A-category two entities, n (for noun, or the semantic equivalent thereof), to be written: A*n*n, where *≠+, and $A*n_a*n_b \neq A*n_b*n_a$; the ordering is arbitrarily fixed so that the first n is equivalent to the subject, and the second to the object, of an action. Each A may be considered as a function (mathematical variety) of two variables (the n's). The n's may also be visualized as occurring in a separate dimension. When they appear in rules, they will be written in the string using the *-notation. When they are irrelevant in rules, they will not be written.

With a convention such as (8) we can describe certain co-occurrence phenomena in fairy tales. For instance, if the villainy in a tale is a kidnapping, the person kidnapped will either be the hero's sister, a priest's or a merchant's daughter, a princess, the hero's bride, etc., but not the king, a dragon, a wicked prince, an old hag, etc. If the villainy is a murder, the victim may be the hero's parents or his father, but not the hero's sister, his bride, an old hag, a wicked prince, etc.

The different kinds of villainy can be enumerated using rules similar to subcategorization rules in syntax and lists similar to lists of lexical items. For example,

$C_V \rightarrow \left\{ \begin{array}{l} C_V \\ C_V \\ . \\ . \\ . \end{array} \right\}$ (C_V is rewritten as either C_{V1} or C_{V2} or ...)

$C_{V1} \rightarrow$ kidnapping, placing under a spell, ...

$C_{V2} \rightarrow$ murder, ...

The restrictions on possible victims can be given by context-restricted phrase-structure rules, that is, phrase structure rules that apply only in some stated context[11]. For example,

$n \rightarrow n_1 \ / \ C_{V1} {}^*X^* \_\_$

$n \rightarrow n_2 \ / \ C_{V2} {}^*Y^* \_\_$

$n_1 \rightarrow$ sister, princess, bride, priest's daughter, ...

$n_2 \rightarrow$ parents, father, ...

Notation: "/" means "in the context." "__" refers to the position of the symbol to the left of the arrow. X and Y are variables. We will continue to use capital letters from the end of the alphabet to indicate variables.

Although context-restricted phrase structure rules will suffice to describe the constituent structure of a simple tale as well as co-occurrence restrictions of the above sort, they will not do to account

for certain other co-occurrence phenomena. For example, in the tale given by (6), the object of I must be the subject of C_v. In other words, Ivan must fight and overcome the same villain who kidnapped his sister. The tale would be "senseless" if Ivan had fought and killed an ogre or an old hag, when a dragon had kidnapped his sister. And it would be either senseless or comic if he fought a dragon and it turned out to be the wrong dragon[12].

It would be possible to describe this phenomenon using context-restricted phrase structure rules, but it cannot be done with rules of any generality at all; for each villain, there would have to be a different rule. The rules might take the following form:

(9) 1. n → n_v / I*X*__
2. n_v → Koscey, dragon, ...
3. n → Koscey / C_v*__*X+Y+I*Z*Koscey
4. n → dragon / C_v*__*X+Y+I*Z*dragon
.
.
.

If there are a hundred possible villains, then with phrase structure rules there must be one hundred rules of roughly the same form to insure that the person who commits the villainy is the same person the hero defeats. This is an unsatisfactory solution; if we use words we can give one simple general rule to describe this phenomenon (as we have done in the previous sentence). If the formal apparatus of context-restricted rules forces us to write one hundred rules where there should be one, then we should seek a more powerful formal device for the description of tales. Transformational rules[13], as developed by Noam Chomsky, provide us with a sufficiently powerful apparatus to describe such phenomena. Using a transformation, we can write the rules of (9) as:

(10) 1. and 2. same as in (9)

3. Structural Description:

C_v*X*Y+Z+I*W*n

1—2—3—4—5—6—7

Structural Change:

1—2—3—4—5—6—7 ⇒ 1—7—3—4—5—6—7

Whichever villain has been chosen by rule 2 is made the subject of the villainous action.

Rules with transformational power can very simply describe otherwise complicated phenomena. In discussing the way in which tales may be combined, Propp points out (83-84) that many complex tales are formed in one of the following ways:

1. One move immediately follows another.

4. A folktale may begin with two villainies, of which the first one may be liquidated completely and the other taken care of later.

5. Two moves may have a common ending.

These are all examples of the same phenomenon: conjunction.

1. has the structure: Plot+Plot

4. has the structure: CS+RS+RS

5. has the structure: CS+CS+RS

All of these may be described by the same transformation:

(11) SD: W+X+Z; W+Y+Z
1—2—3—4—5—6

where: X and Y are the same constituent, whether Plot, CS, or RS.

SC: 1—2—3—4—5—6 ⇒ 4—2—5—6

(This transformation takes two moves into one complex move. The two moves are separated by a semi-colon.)

Where X and Y not only form the same constituent, but are identical, we have a case of repetition.

Another transformation can describe the case where a tale is interrupted by an episode, or a series of episodes, and then continues. Episodes usually occur after the donor sequence and before the resolution. We can account for them by one rule change and one transformation. Introduce an optional episode marker EP in (6)-4.

4. RS → DS+(EP)+R+K+Rew

(12) SD: Plot; X+EP+Y
1 — 2—3 —4

SC: 1—2—3—4 ⇒ 2—1—4

Actually, this is somewhat simplified. Usually, there is the added restriction that the hero of the episode must be the hero of the main plot. In order to incorporate this information in the description, we must be able to define the hero. Perhaps the best way to do this is to consider the hero as the person who defeats the villain. Then, the hero will be X in the expression: I*X*Y. The rule for the embedding of an episode would then be:

(13) SD: W+I*X*Y+Z; Q+EP+T+I*U*V+S
1 — 2—3 — 4

where: X=U

SC: 1—2—3—4 ⇒ 2—1—4

Not only can transformation rules describe many features of the composition of tales simple, but they can in some cases <u>explain</u> certain of our intuitions about tales. As Propp points out, there is some confusion between the elements of the donor sequence and the elements of the resolution. Propp discusses the following case (p. 60):

> Ivan asks the witch for a horse. She proposes that he select the best of a herd of colts. He chooses accurately and receives the horse. Here, the action at the witch's house is the test of the hero by the donor. This is later followed by the receipt of a magical agent. In another tale, on the other hand, we see that the hero desires to wed the daughter of the water spirit, who requires the hero to choose his bride from among twelve identical maidens. Can this case, as well, be defined as the donor's test? It is clear that, in spite of the identical quality of the actions, we are confronted with a completely different element, namely a difficult task connected with match-making. An assimilation of one form by another has taken place.

In discussing the A-category, H, Propp comments (p. 47):

> This form needs to be distinguished from the struggle with a villainous donor. These two forms can be recognized and contrasted according to the effects they produce. If the hero obtains an agent for the purpose of further seeking, as the result of combat with a villainous character, this would be element D. We would designate as element H a situation whereby the hero would receive, as the result of combat, the very object of quest for which he was dispatched.

The facts are essentially these: the same things can happen in a donor sequence as can happen in the resolution of a simple tale. The person who is the villain in the tale is the donor in the corresponding donor sequence. The donor sequences in which there is a struggle with a villainous donor correspond exactly to the resolutions in which there is an H-I (struggle-victory) sequence. The donor sequences in which the donor asks a riddle or proposes a difficult task correspond exactly to those resolutions in which the villain sets a difficult task. Moreover, the same characters who can serve as villains can also serve as donors. All the evidence leads us to suspect that the donor sequence is really part of another simple plot embedded in a given simple plot. We can describe this embedding by altering some of the rules given in (6) and adding a transformation.

(6') 4. $RS \rightarrow (DS)+(EP)+R+K+Rew$

Eliminate rule 5

7. $R \rightarrow \left\{\begin{matrix} H+I \\ M+N \end{matrix}\right\}$

6. $C \rightarrow \left\{\begin{matrix} C_V \\ C_L \end{matrix}\right\}$

(M: propose difficult task
N: complete difficult task
C_L: Lack)

(The villain may now also be described by n in the expression: M*n*X.)

(14) SD: $C_L+RS;\ X+DS+Y$
$1-2-3-4-5$

SC: $1-2-3-4-5 \Rightarrow 3-2-5$

Or, if you include the condition that the hero must be the same in both:

(15) SD: $C_L+X+\left\{\begin{matrix} I \\ N \end{matrix}\right\}*Y*Z+W;\ T+DS+U+\left\{\begin{matrix} I \\ N \end{matrix}\right\}*V*Q+S$
$1-2-3-4-5$

SC: $1-2-3-4-5 \Rightarrow 3-2-5$

Notice that by introducing the donor sequence in this way, we have eliminated the necessity for listing the possible entries for A-categories D, E, F, and G, and for duplicating the list of possible villains as the list of possible donors. More important, we have explained the relationship between the elements of the donor sequence and the elements of the resolution. In addition we have explained such a traditional puzzle as why Baba Jaga appears as both donor and villain. Baba Jaga is listed among the villains, but villains in embedded donor sequences are called "donors."

IV

Any list of rules in a model for a "grammar" of fairy tales must at this time be both tentative and incomplete -- at least because insufficient research has been done and, perhaps, for some essential reason, i.e., it may not be possible to construct such a model. We doubt that there is any essential roadblock in constructing such a model and we take the position that accurate models of some sort are possible not only for fairy tales but for all kinds of coherent human discourse.

The following list of rules is a collection of those mentioned in the preceding section. A large part of Propp's work is accounted for in these rules. We have omitted several of Propp's functions because we were not satisfied with them, but could not come up with a better solution.

The rules are ordered and form two levels of analysis. On the first level, phrase structure rules are sufficient. The output of this level is the set of simple plots. Complex plots are formed in the second level, where transformation rules are introduced.

PHRASE STRUCTURE:

1. Plot → CS+RS
2. RS → (DS)+(EP)+R+K+Rew
3. CS → (HS)+C+L
4. R → $\begin{Bmatrix} H+I \\ M+N \end{Bmatrix}$

5. HS → Int+Vio

6. Vio → $\left\{\begin{matrix}\text{WVio}\\ \text{DEC+SUB}\end{matrix}\right\}$

7. C → $\left\{\begin{matrix}C_V\\ C_L\end{matrix}\right\}$

(WVio: willful violation
DEC: deception by villain
SUB: submission of the hero)

At this point apply convention (8) (p. 138 above).

---

8. C_V → $\left\{\begin{matrix}C_{V1}\\ C_{V2}\\ .\\ .\end{matrix}\right\}$

9. C_L → $\left\{\begin{matrix}C_{L1}\\ C_{L2}\\ .\\ .\end{matrix}\right\}$

10. n → n_v / $\left\{\begin{matrix}I*X*\_\_\\ M*\_\_*Y\end{matrix}\right\}$

11. n → n_h / $\left\{\begin{matrix}I*\_\_*X\\ N\end{matrix}\right\}$

12. n → n_{vm1} / C_V *X*__

13. n → n_{vm2} / C_V *X*__

Other subcategorizations.

Lists:

14. n_v → Baba Jaga, dragon, Koscey, ...

15. n_h → Ivan, young prince, Sucenko, young peasant lad, ...

16. n_{vm1} → sister, princess, bride, ...

17. n_{vm2} → parents, father, ...

18. C_{V1} → kidnapping, placing under a spell, ...

19. C_{V2} → murder, ...

20. H → engage in battle, ...

21. I → defeat, kill, ...

22. M → set difficult task

23. N → accomplish difficult task

Other lists

TRANSFORMATIONS:

1. Villain Placement: SD: C_V*X*Y+Z+$\begin{Bmatrix} I*W \\ {}_1M_1 \end{Bmatrix}$*n$\begin{Bmatrix} +V \\ {}_1*U+T_1 \end{Bmatrix}$

1 —2— 3 —4— 5

SC: 1—2—3—4—5 ⇒ 1—4—3—4—5

Other such co-occurrence rules.

2. Conjunction: SD: U+X+V; U+Y+V where: X and Y are both either Plot, CS, RS, R or C
1—2—3— 4—5—6

SC: 1—2—3—4—5—6 ⇒ 4—2—5—6

3. EP-EMBEDDING: SD: W+$\begin{Bmatrix} I \\ N \end{Bmatrix}$*X*Y+Z; Q+EP+T+$\begin{Bmatrix} I \\ N \end{Bmatrix}$*U*V+S where: X=U

1 — 2—3 — 4

SC: 1—2—3—4 ⇒ 2—1—4

4. DS-EMBEDDING: SD: C_L+X+$\begin{Bmatrix} I \\ N \end{Bmatrix}$*Y*Z+W; T+DS+U+$\begin{Bmatrix} I \\ N \end{Bmatrix}$*V*Q+S where: Y=V

1 — 2 — 3—4 — 5

SC: 1—2—3—4—5 ⇒ 3—2—5

In order to convert the output of the transformational level into a connected string of sentences, two sets of lower level rules will be required. First, there will be sentence-forming rules that will take abstract expressions like kidnap*Koscey*princess and convert them into sentences of a particular language, say English or Russian. What these rules will look like is anyone's guess. At the present time it is not even certain that such a goal is possible. Assuming, however, that such rules can be formulated, they will have to be followed by a set of rules to assure that the narrative will be connected. It is not at all clear how such a model will mesh with the existing model for the generation of sentences.

V

If we add to the rules of the previous section the additional co-occurrence restrictions and lists indicated and the required sentence

formation and connectedness rules, the resulting "grammar" will generate strings of sentences that we can recognize as fairy tales. Complicated as many of these will be -- what with conjoined and embedded sequences -- they will still be rather skimpy, almost skeletal tales. Most of the colorful details, stylistic features, and intricate convolutions of plot will still be missing. Until someone finds the formal apparatus to describe them, the study of fairy tale structure will be far from complete. And, even though we have succeeded in describing on two levels the structure of a very small subset of the set of discourses, it is by no means clear that we can extend this sort of description non-trivially to other kinds of discourses -- scholarly essays, political speeches, and conversations, to name a few. In short this work can at best be considered the barest beginning on an intriguing route of investigation.

Nevertheless, we can say with some certainty that people do construct some discourses in much the same way that they construct sentences. If people have in their minds a "discourse grammar," it is certain that large segments of that grammar will be shared by people speaking many different languages in many scattered lands. Most Western Europeans, for example, will understand tales from other Western European countries. It seems equally certain, though, that all people do not share the same discourse grammar; at least it seems certain on the basis of folktale evidence. Most Europeans and Americans of European ancestry find the great body of American Indian tales largely incomprehensible -- often both senseless and structureless.

If it is so that not all people put thoughts together in the same way to form discourses, then we shall have an interesting new phenomenon on our hands. Most of us tacitly assume that people who speak languages different from ours merely talk differently -- and may possibly hold different world views and different cultural values -- but we assume that they put their thoughts together pretty much like everyone else. This may be approximately true for most Americans and Western Europeans, but it may well not hold true for everyone.

It may even be the case that people who speak the same language and are raised in the same cultural circumstances will have slightly different discourse grammars, just as they have slightly different grammars of whatever language they speak. This may explain many cases where one just does not "get the point" of a story a friend tells. Observations such as these may be interesting insights into the nature of language, or they may be merely errant speculations. But before we can determine which they are, we must achieve a precise formulation of a theory of discourse.

REFERENCES

1. Afanas'ev, A.N. Russian Fairy Tales. Translation by Norman Guterman. Notes by Roman Jakobson. New York: 1945.
2. Chomsky, N. Syntactic Structures. The Hague: Mouton, 1957.
3. Chomsky, N. Three Models for the Description of Language. IRE Trans. on Inform. Theory, 1956, IT-2, 113-124.
4. Chomsky, N. On the Notion "Rule of Grammar." In R. Jakobson (Ed.), Structure of Language and its Mathematical Aspects. Proc. 12th Symp. in App. Math. Providence, R.I.: American Mathematical Society, 1961, pp. 6-24.
5. Chomsky, N. Aspects of the Theory of Syntax. Cambridge: MIT, 1965.
6. Newmark, L. and Bloomfield, M.W. A Linguistic Introduction to the History of English. New York: Knopf, 1963.
7. Propp, V. Morphology of the Folktale. Translation by Laurence Scott. Edited by Svatava Pirkova-Jakobson. Supplement to International Jour. of Amer. Linguistics. Baltimore: 1958.

FOOTNOTES

1. See Newmark and Bloomfield, 70-71.

2. See Propp, 11-12.

3. " " 18.

4. " " 19.

5. " " 19-20.

6. " " 20-21.

7. Since human memory is finite, we must require our model to be finite. Whatever finite device we choose must however be capable of generating an infinite number of tales. To demonstrate that there is no formal bound on the number of possible tales, it is sufficient for us to state an algorithm capable of producing an infinite set of tales. We can start with the following tale, T_1:

> $Irving_1$ grew to manhood in the hamlet of Schlimmielsburg. One day a dragon entered the village and threatened to destroy it. Though the other inhabitants fled, $Irving_1$ entered into hand-to-claw combat with the dragon and overcame him. $Irving_1$ was celebrated as a hero and given the mayor's daughter's hand in marriage.

The algorithm would go like this: Given T_i, form T_{i+1} by adding to T_i the sentence "She bore him a son, $Irving_{i+1}$." and then adding S_1 with each occurrence of $Irving_1$ replaced by $Irving_{i+1}$.

The set of stories produced in this way would be enumerably infinite. Stories formed by devices similar to this form an important part of the folk tales and epic songs of many cultures. There are seldom, if ever, any formal bounds on the lengths of such tales; they are limited only by the endurance and imagination of the tale teller.

See Chomsky (1957), 23.

8. See Chomsky (1957), 21.

9. See Chomsky (1956).

10. See Chomsky (1956).

11. These rules can be written simply for the fairly simple example considered here. If, however, further study should reveal that some subcategories require cross-classification (are linearly independent) then PS rules will be replaced by complex symbol rules. See Chomsky, [1965], Ch. 2.

12. This is the basis of a time-honored comedy routine. We are reminded of the Abbott and Costello bit, in which Costello bags a supposed bank robber only to discover he's caught the police chief or bank president. Such comic episodes are often based on deviance from an established pattern. We are concerned here with the established pattern.

13. Transformations are, roughly, combinations of permutations, additions, deletions, and substitutions. For a precise formulation, see Chomsky (1961).