PETER KÜMMEL

AN ALGORITHM OF LIMITED SYNTAX BASED ON LANGUAGE UNIVERSALS

Abstract.

Hundreds of differentiated syntactic rules as well as all kinds of phrase structures and conventions are dispersed over more than fifty still existing natural language systems. On the contrary meanings, represented by nouns, verbs and adjectives are internationally identical. Only the way they are carried and expressed by various national languages differs. Thus a fundamental division into two parts represents the main feature of the ALGORITHM OF LIMITED SYNTAX (ALS). This ALS-Division of natural language systems provides two groups of words:

a) words representing nouns, verbs and adjectives as so called « content words » or more easily nominated as *RADICALS* and

b) words expressing syntactic features and functions, the *SYNTAX-PARTICLES.*

While the human treasure of knowledge in form of meanings expressed by content-words is in no way curtailed by ALS-Rules, the vast jungle of still existing syntactical features and syntagms will be drastically cooked down to the essentials. Thus the ALS-Rules comprehend several laws to limit syntagms, which are made clear by algorithms or verbal definitions. Still the limitations enforced by ALS-Rules are based on language universals and thus keep all information-expressions and -performances in the scope of a natural language system.

1.0. *Aims and the unconventional character of approach.*

Significance and necessity to solve the problems concerning Automatic Translation between different national languages as well as fact retrieval, justifies all possible approaches. Worldwide pessimism (M.

L. MINSKY, 1968, p. 24 f) relating to this enterprise ten to twenty years ago (N. CHOMSKY, 1965, p. 163) has been in no way without foundation (Y. BAR-HILLEL, 1940, p. 271 and G. SALTON, 1968, p. 156 f). A *Formalization of natural language systems* seems for the first step unavoidable, if the three fundamental goals:

1. AUTOMATIC TRANSLATION

2. AUTOMATIC FACT RETRIEVAL and

3. ARCHIVAL HARD STORAGE, a once and for good storage of human information in form of facts, are aimed at.

In the thirties L. BLOOMFIELD published a very significant idea (1935, p. 208) of dividing natural language systems among others into those of

a) agglutinated morphologies and
b) those of isolated morphologies.

While English, French, German and Russian belong to systems of agglutinated morphologies – the morphological structures are glued together – Chinese, ideographic-, pictographic- and gesture language systems bear an isolated one. When dealing especially with phenomena of isolated morphologies, two points become very clear:

1. The multitude of syntagms i.e. syntactic features and patterns coming with agglutinated systems are considerably reduced to the very important ones.

2. The mostly visually to perceive expression units – permanent as well as acute – (grams and gestures), are highly content related, thus helping to look behind the scenery of language expressions into the material of content-phenomena and semantic relations.

Up to fifty or more existing different language systems can be considered as morphologically and syntactically differentiated expression – and carrier – systems for only one and the very same information treasure, the treasure of human knowledge. So it is necessary to unify expression – and syntax – habits. In order to unify and subsequently limit especially the number of existing syntagms, it pays off to be very careful when making artificial curtailments. The best way still is to lean on natural prototypes, which lead back to the Bloomfieldian division (L. BLOOMFIELD, 1935, p. 208):

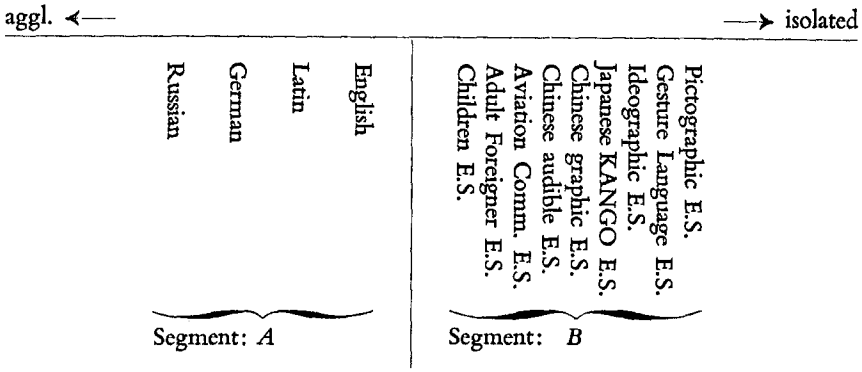


Fig. 1. *Scale of Language Carrier-Systems: Agglutinative to Isolated*

Consequently it becomes convenient to disregard segment *A* of Fig. 1 entirely and look more into details of segment *B*. Syntagms, limited down by their nature in language expression systems of segment *B* are easier to pinpoint. The vast jungle of syntagms in segment *A* hinders the extraction of those syntagms which are of most significance concerning language universals. They are too much covered and hidden by conventional balast of low-value syntagms (Fig. 2).

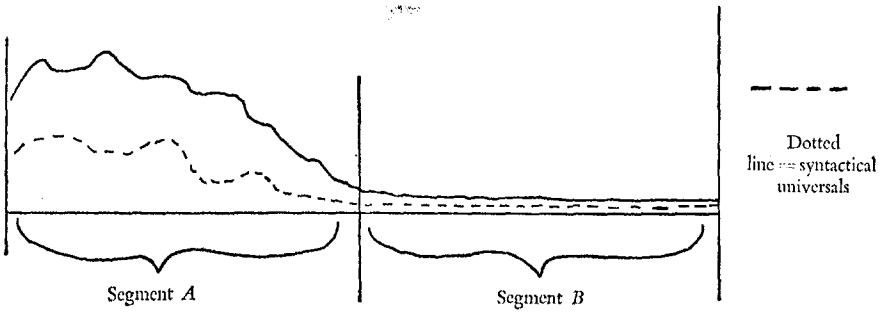


Fig. 2.

The number of possible syntactic algorithms to construct sentences in all existing natural languages is finite, but still at the present state " uncontrollable " and thus not computable (G. SALTON, 1968, p. 191). Obeying this fact of nature, efforts led to the conclusion, not to operate on sentences of natural language as they exist now, but to form a rule and a recipe to conceive sentences, which are:

1. natural in character and
2. easy to formalize and compute.

Even if limited in syntactic patterns, natural language systems are easy to employ because of being available and programmed already in the brains of human beings. Interpreting the criteria for natural language systems (J. H. GREENBERG, 1968, p. 110 ff), it is trivial to realize that certain systems, up to now mostly out of regard, deliver contributions of syntactical features highly valuable for the whole enterprise. Those syntactical features of limited but natural expression systems are mainly extracted from language systems of isolated morphology as shown in Fig. 3, e to m. But the entire operation to limit syntagms and form algorithms must be stopped by one certain point, which prevents it from crossing the boundaries from natural to artificial language systems (Fig. 3). This boundary has been easily pinpointed by following the rules of a work published one year ago. In this publication a distinct separation between natural and artificial language systems has been made per definitionem (P. KÜMMEL, p. 86):

Definition: NATURAL LANGUAGE

A language is natural in its character, if

a) the number of using individuals is larger or equals one million heads and

b) the period of use is larger or equals one life span of a using individual

Natural Languages			Artificial Languages	
a	Russian	m	Japanese KANGO E.S.	p various progr. Languages
b	German	l	Pictographic E.S.	o Ancient Greek
c	French	k	Gesture Language E.S.	n Latin
d	English	j	Ideographic E.S.	boundary
e	Children E.S.	i	Chinese graphic E.S.	
f	Adult Foreigner E.S.	h	Chinese audible E.S.	
g	Aviation Comm. E.S.			

Fig. 3.

The remainder of language systems can be considered artificial. Artificial also insofar as a few native language systems, presently only used by some thousand individuals, have a too limited number of users and thus lack the natural momentum. Within the scope of natural languages (Fig. 3, e to m) a rather wide gap exists between d and e. The closer a language system is located to the boundary between natural and artificial system (Fig. 3, m/n), the easier it becomes to formalize it, as the balast of conventional syntactic rules fades away. So it paid off quite heavily to analyze language carrier systems especially from e to m (Fig. 3). It is a pity that so few linguists have involved themselves with language systems of isolated morphologies as shown in e to m in Fig. 3. This fact explains why the author of this publication has to refer to eight of his own publications in the reference literature.

The language systems in Fig. 3, e to m are natural in character and still after the Bloomfieldian definition of isolated morphology. Thus they are easier to formalize. They represent:

e) Children Expression Systems

f) Adult Foreigner Expression Systems, as found among tourists or guest-labourers with a very limited vocabulary treasure and almost no knowledge of syntactical patterns. Habits practiced by adults who are forced to maintain dialogs in foreign languages without preparatory study. Language of people who stay only a limited time in foreign countries.

g) Aviation Control Communication Language as used between the pilot and tower personal.

h) Chinese for the audible perception.

i) Chinese in written form for the visual perception.

j) Ideography and ideographic expression systems such as Hieroglyphs, etc.

k) Gesture Expression Systems as for example deaf-mute, indian and nomadic gestures for the acute visual perception.

l) Pictographic Expression Systems as found in traffic signs, etc.

m) Japanese KANGO Expression Systems, a form of ideogram sequences referring back to Chinese.

A conglomerate of universals in syntactical features to be found especially in the preceding language systems e to m but also among those of a to d (Fig. 3) is filtered and concentrated, to retain syntactic phenomena, characterized by the highest frequency of occurrence (Mc. D. Neill, 1971, p. 40). Thus an ALGORITHM OF LIMITED SYNTAX (ALS) in the form of several fundamental syntagms is gained, which pre-

serves a full scale consolidation of the natural momentum and enables the user to feel familiar at once.

The preceding details led to a somewhat unconventional approach characterized by the following three features:

1. In order to apply a few and most essential syntactical algorithms for natural languages, the content metric of one meaning is enlarged. SHANNON/WEAVER's information measurement (entropy/bit) has been extended by evaluating eight content criteria or so called sub-criteria of one information unit, one meaning or one *RADICAL* (P. KÜMMEL, 1972 a, p. 64 ff): 1. Identity, 2. Truth, 3. *RADICAL*-Value, 4. Age, 5. Common Frequency, 6. Relative Frequency, 7. Priority and 8. the Interlogging Value.

2. A conglomerate of syntactic universals found especially in content-related natural language systems is cooked down to the essentials in form of a few syntactical algorithms with highest frequency of occurrence. This conglomerate in cooperation with the newly conceived content evaluation (1) is called: ALGORITHM OF LIMITED SYNTAX (ALS).

3. The amalgamation of the preceding features (1) and (2) in form of the very ALS serves for a case study. This case study involves an execution of Archival Hard Storage, Automatic Fact Retrieval and Automatic Translation of Natural Languages. It covers a small molecular scope of human knowledge centered around the meaning " Sand Glass " and the " father – son – relation ". Those artificial question answering systems on structural and functional phenomena of the domain " Sand-Glass " as well as the " father – son – relation " can be considered a recipe for automated handling of the remaining human knowledge treasure on earth. Concessions to be made appear in the form of stereotype syntactical patterns, which affect the esthetic momentum but not the transmission flux of information. A natural text to be computed is conceived after application of the easy to learn and easy to limit ALS-Rules, which makes it natural in character. Present day used texts of Natural Languages must be rewritten.

2.0. *A RADICAL, the newly conceived content unit and its sub-criteria in the ALS-system.*

In 1928 Hartley originated the idea of demonstrating the canal capacity of a system mathematically. He put the logarithm of the sum

concerning all possible situations equal to the information capacity (R. v. L. HARTLEY, 1928, p. 535 f). Later this was reversed by Shannon and Weaver. In respect to the probability of choice Weaver stated: " The actual expression for information is " (W. WEAVER, 1964, p. 14):

$$H = - \Sigma p_i \; log \; p_i$$

Expressed by the logarithm dualis in bits Shannon declared the probability of occurrence to be high, if the information value is small. In case the probability of occurrence reaches 1 the information equals 0.

$$I \; (a_i) = - \; 1d = 0 \; bit$$

Consequently if $p = 0$, it will cause the information to become indefinite (C. E. SHANNON, 1964, p. 63):

$$I \; (a_i) = - \; 1d \; 0 = \infty \; bit$$

This calculation, which is analogous to the entropy evaluation of thermostatistical data was used (1963) by Beer to express the dimensions of information values by bits. He associates several different inscriptions on train windows with the corresponding information value (S. BEER, 1963, p. 65):

Do not lean out of the window	137.75 *bits*
E pericoloso sporgersi	104.50 "
Lisfarligt att luta sig ut	128.25 "
Ne pas se pencher aus dehors	128.25 "
Nicht hinauslehnen	85.50 "

Since that time, developments have shown that a strict limitation to only this kind of content evaluation might lead into a blind alley. Bit evaluations are sufficient under certain circumstances to measure carrier systems and signs as their units, but they fall absolutely short in the determination of contents in the form of information. A newly born scientific fact or an invention represents one meaning and can be named with one sentence, one word or even one sign. This meaning stands for an information unit and is criticized by more than the value " improbability of occurrence ", which might come close only to No 6 of the eight content criteria.

An information unit, represented by an abbreviated noun, verb, or adjective in lemma - or better said stem - form, is called *RADICAL*. A *RADICAL* is the main and only component in ALS - sentences besides the particles. The *PARTICLES* carry exclusively syntactical functions. A *RADICAL* can be subdivided into eight further subcriteria, which also are to be enumerated.

Quite well known among Orientalists and about 1 billion asiatic people are the so called *RADICALS*, after which composed Chinese Characters can be classified (P. KÜMMEL, 1966, p. 110 f). The system of *RADICALS* delivers for the information theory as well as Semantics of natural languages a highly valued contribution to form sub-criteria of one meaning. In other words they represent the interlacing-, entangling- or involving-value of one meaning to others. With the aid of *RADICALS* several sub-criteria of one meaning can be formalized and enumerated to a 100% degree of precision. Below, five examples of Chinese *RADICALS* are given in connection with several compositive characters, in which the *RADICAL* represents one component:

RADICALS:		Composed Characters:							
革	*leather* =	靴	*shoe,*	鞄	*suitcase,*	鞍	*saddle,*	鞭	*whip*
木	*tree* =	机	*desk,*	橋	*bridge,*	械	*machine,*	柱	*shelf*
石	*stone* =	岩	*rock,*	砂	*sand,*	碁	*Go-Stone (a game)*		
糸	*thread* =	結	*to connect,*	紙	*paper,*	続	*to continue,*		
		細	*narrow,*	織	*to weave*				
貝	*mussel* =	賃	*to rent,*	賃	*wage,*	貰	*receive,*	賣	*to sell,*
		買	*to buy,*	員	*member/person*				

There are already existing expression systems in Natural Languages, which come close to ALS-habits. Besides the Chinese expression system for example the Japanese KANGO-System is known. 漢語 means literally " Chinese words ". In a Japanese text theoretically all particles have to be eliminated, in order to generate a KANGO-Text. For example: The sentence: *The car runs fast* can be written:

車が速く走ります。

1 2 3

If 1 the nominative particle, 2 the adjective particle and 3 the verb-suffix are omitted, it results in 車速走, out of which still the same meaning is understood. Examples for used KANGO-Expressions in Japanese are: 駐車難 *difficult to park*, or: 決速電車 *an electrical express train*. Both KANGO-Expressions were conceived after the fifties.

Finally, the use of *RADICALS* with a limited number of *PARTICLES* comes not only close to CHINESE and Japanese KANGO-Texts, but also offers the opportunity to subdivide one *RADICAL*-into several content-criteria.

The eight sub-criteria of a *RADICAL*:

1. *IDENTITY*

A new information unit can be interpreted and integrated here and there. It can be filed here and there. The identity is implemented by the *RADICAL*-Number (*Deuterziffer*) (P. KÜMMEL, 1972 b, p. 483 f). The first meaning in the human treasure of knowledge receives the number 1. The *RADICAL*-Number is implemented at a graph's vertex (Fig. 4) in the way of decimal fractures used for decimal classifications. The *RADICAL* represents a vertex

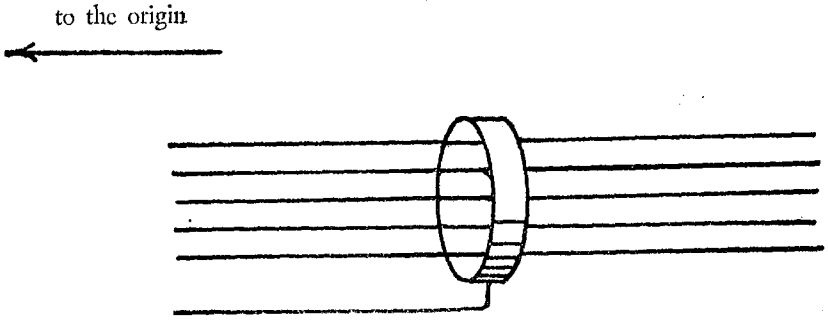


Fig. 4.

in a graph which also can be shown as a loop with connection strings like a cable-tree.

2. *TRUTH*

If a new information represents a perpetuum-mobile or a product of fantasy, it has to be marked false. The truth is implemented into the treasure of knowledge graph by even numbers making use of the *IDENTITY*-Numbers under 1. If the truth is not existing and an implementation has to be done, odd *RADICAL*-Numbers are provided.

3. *RADICAL-VALUE*

It is more or less related to existing information, which makes it attached as a green bud to thicker or thinner branches of the information graph (Fig. 5). The *RADICAL-VALUE* is also to compare with the

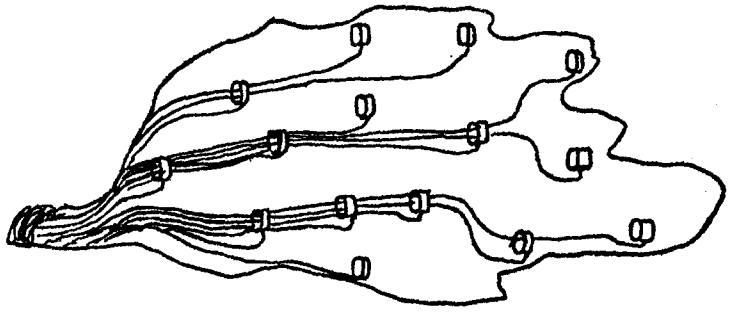


Fig. 5.

distribution ability of a vertex (P. KÜMMEL, 1969, p. 64 ff) and thus is identical to the number of strings in the vertex loop (Fig. 6). The *RADICAL-VALUE* is written in brackets at the *RADICAL*-Loop or the vertex. For more detail see P. KÜMMEL (1972 b, p. 481 f) and P. KÜMMEL (1972a, p. 36 f).

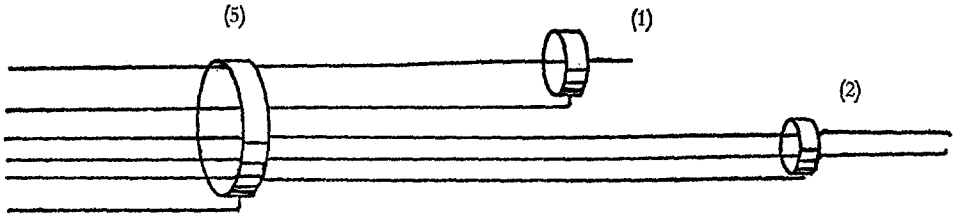


Fig. 6.

4. *AGE*

If an information unit or *RADICAL* turns out to be not new, the exact age has to be determined. The implementation of age is done by chronological numbers, starting with 1 for the oldest and furnishing the highest in use for the latest *RADICAL*. Secondly the age of a *RADICAL* may be determined in relation to a known time-calendar. So it receives a birth-date.

5. *COMMON FREQUENCY*

If a *RADICAL* is already known, it can be classified by its common frequency since its birth. The *COMMON FREQUENCY* is implemented by numbers indicating the occurrences of counts (P. KÜMMEL, 1968 b, p. 242 ff).

6. *RELATIVE FREQUENCY*

In comparison to others, one information occurs for practical use oftener or not. The information unit with the highest relative occurrence is designated with the number one (1). Another of half of that rate by 1/2 and so on (P. KÜMMEL, 1968 b, p. 242 ff). The Relative frequency substitutes more or less Shannon/Weaver's Bit-Metric.

7. *PRIORITY*

A larger or smaller importance compared to other meanings is indicated by a number which states the surface tension of a *RADICAL*-Graph in those places, where the surface is grown unregular (Fig. 7). The vertex *A* is located close to a surface area where high tension exists (see arrows!). It owns a higher *PRIORITY*-Value than the knots *B* and *C*. The *PRIORITY*-Value is measured in percent. If the tension on the *RADICAL*-Cluster

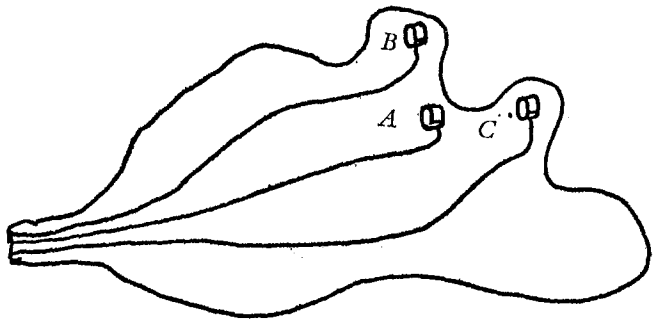


Fig. 7.

(see arrows!) is strong enough, the *PRIORITY*-Value can reach up to 100% (P. KÜMMEL, 1972 b, p. 483).

8. *INTERLOGGING-VALUE*

An individual associates with a certain meaning (*RADICAL*) another meaning, which lies far away from the Cluster-Branches of the first concern (Fig. 8). Thus this association cannot run along neighbouring branches, but jumps to a *RADICAL*-Loop far away via the dotted lines

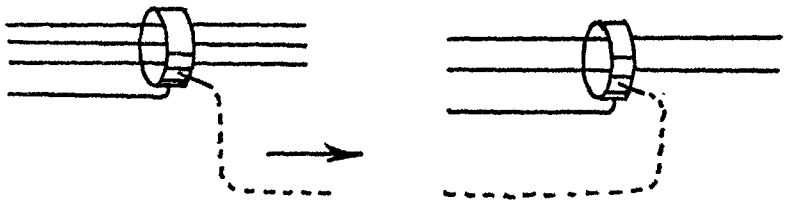


Fig. 8.

(Fig. 8). The extent of the *INTERLOGGING*-Value is given by the distance of the dotted line, to be measured by the differences of the Identity-Value in the Radical-Cluster. Secondly the *INTERLOGGING*-Value depends on the number of interlogging- and distance-jumps via dotted lines, in other words by the number of dotted lines.

Definition: *RADICAL*

A *RADICAL* represents a meaning, which can be subdivided into so called sub-criteria or *RADICAL*-Criteria. The *RADICAL* represents the counter-part to the *PARTICLE* in ALS-Texts. It carries the meanings of nouns, verbs and adjectives. For automatic translation purposes a *RADICAL* is translated exclusively into different national languages.

3.0. *The PARTICLES, a limited number of syntagms permitted in the ALS-System.*

The human knowledge is to be considered as a huge cluster of nowadays not less than several million *RADICALS*. But a language, functioning as a carrier system for information and thus single *RADICALS*, is always composed of syntagms too. These syntagms finally combine sequences of *RADICALS* to sentences of Natural Language Systems. They can be reduced to a minimum of units, but may not be neglected entirely. A very carefully done and delicate process of limiting down the syntagms resulted in a concentrated stockpile of abbreviated *PARTICLES* integrated into the ALS-Rules. The *PARTICLES* are abstract in their character and each embodies a motion in the branches of the *RADICAL*-Cluster. Consequently the use of a *PARTICLE* releases an electrical process within an associative hardware storage. The *PARTICLES* in isolated form do not supply much information. They become useful only when appearing in connection with *RADICALS*. As some examples of elementary *PARTICLES* can be considered:

1. /(quest.)/ for question and
2. /(pret.)/ for preteritum or past tense.

Definition: *PARTICLE*

In ALS-Texts the *PARTICLE* represents the counterpart to the *RADICAL*. A *PARTICLE* serves as a fundamental syntagm exclusively syntactical functions. The *PARTICLE* is considered a concentrate of internationally universal syntactic rules. For automatic translation purposes the *PARTICLE* is not translated, thus appearing in internationally understood abbreviated forms of Latin or Greek.

4.0. *Fundamental rules of the algorithm of limited syntax divided into single paragraphs.*

In order to use ALS, conventional sentences must be transformed to heteronoms with the maximum size of a fivefold-heteronom. Different from other approaches to the same project, where at least 172 syntactic features must be taken into consideration (T. WINOGRAD, 1972, p. 173 f), ALS limits the formalism by permitting only stereotype patterns. These patterns include contextual rules within the nominal-sequences of up to five nouns. Only a restricted number of *PARTICLES* within the fivefold-heteronom is permissible. ALS-Formalisms are based on the "*Grundregeln der Formalgrammatik*" (P. KÜMMEL, 1972b, p. 485) as well as on details described in the "*Deuterfolgen-Grammatik*" (P. KÜMMEL, 1972a, p. 76 ff and 83 f). The main rules of ALS are classified in paragraphs:

§ 1 All expressions of Natural Language must be divided into *RADICALS* embodying meanings and *PARTICLES* representing syntactical function.

§ 2 Each *RADICAL* has to appear in sentences with an abbreviated morphology, in stem-pattern or lemma-form.

§ 3 All expressions must be cast into sequences of *RADICALS* similar to heteronoms. Example: *glove* = mononom, *Handschuh* = binom in form of a heteronom.

§ 4 Each of these heteronoms may not exceed the size of a fivefoldnomen. In other words, a sentence may not include more than five *RADICALS*. Example of a maximum: /train/terminal/station-master/office /telephone/.

§ 5 Each of these heteronoms according to § 4 is to carry one meaning in form of an information unit.

§ 6 The meaning of a sentence in the form of a heteronom is always to be taken out of the last *RADICAL*, while the preceding ones only serve for more detailed specifications. Example: /information/bank/means bank or accumulation place. But /bank/information/means news from or about banks.

§ 7 Each meaning according to § 5 has to be analyzed by the eight content criteria (P. KÜMMEL, 1972a, p. 69 f) and (P. KÜMMEL, 1973b, p. 482 ff).
Additionally they must be translatable into all existing natural languages.

§ 8 Each meaning or *RADICAL* according to § 7 must be furnished with a national language identification number for example 1-10 belonging

to the most relevant national language systems on earth (see 5.3 below!). Thus for translation purposes the adequate expression sequences belonging to each identification number can be provided.

§ 9 The heteronoms according to § 4 exist of conventional nouns, as well as substantivated verbs and substantivated adjectives as nouns. Example: ... *gives the money* = /moneytransfer/ or *sunny day* = /sunday/ etc.

§ 10 All *RADICALS* according to 1 and 3 should be possibly limited to their lemma patterns respectively shortest understandable form. Example: *Automobile* = *auto* or *television set* = *TV – set.*

§ 11 Each *RADICAL* in lemma-pattern should not surpass the length of a certain sequence number of letters, which is stated here with 20 letters.

§ 12 Each *RADICAL* within a heteronomic sentence must be written within vertical lines. Example: /father/.

§ 13 Longer sentences of conventional text can be reduced to several heteronoms by rewriting. As an example see the text of *Newsweek* below:

§ 14 According to the chainstitch-method (P. KÜMMEL, 1972a, p. 77 and 85) as well as recursive functions (P. KÜMMEL, 1972b, p. 485, § 4) one *RADICAL* of the preceding heteronom is chained into the following one. Example: *The father living in London gives the son studying in Paris money.* 1. A threefold heteronom: /father/London/living/, 2. A threefold heteronom: /father/son/moneytransfer/ and 3. Another threefold heteronom: /son/Paris/study/.

§ 15 Heteronoms of concrete value end with a concrete noun or *RADICAL.* Example: /motor-car/door/. Heteronoms with an abstract value end with an abstract noun. Example: /motor-car/delivery/.

§ 16 Necessary *PARTICLES* expressing for example time, questions, etc., must not be placed behind the last *RADICAL* of the heteronom. Example: *The father gave the money*: = /father/(pret.)/moneytransfer/. 2. *Does the train leave at 12 : 30 h*? = /train/12 : 20 h/ (quest.)/depart/.

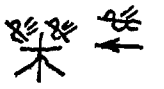
§ 17 *PARTICLES* have to be written in brackets. Example see § 16!

§ 18 Serving the purpose of automatic translation all *PARTICLES* must be expressed in Latin, Greek or an internationally known abbreviated letter sequence, because *PARTICLES* will not be translatable in the ALS-System.

§ 19 The pronoun, which normally represents a noun, always has to be replaced by the originally substituted noun. Example: *Little Joe does not say: " I want to eat "*, but: /Joe/want/eat/. Or, *the pilot does not say: " yes "* or *" I acknowledge "*, but: /Cessna/niner/three/victor/affirmative/.

§ 20 For each *RADICAL* to be used with ALS-Systems a fully content-related expression unit in form of a pictogram has to be furnished.

This must be available in an adjacent archive. Thus ambiguity by carrier functions of different languages as well as homo-and synonymities can be eliminated. Even for the conception of content related abstracts pictograms are possible, see (P. KÜMMEL, 1972a, p. 53 ff). The meanings stay constant in time (P. KÜMMEL, 1967b, p. 42 ff) and (P. KÜMMEL, 1968*a*, p. 52 f). Pictograms of abstracts must be conceived with the fourth dimension " time " in them, realized by an arrow or a dotted line for deplacements and movements. For example in China and Japan the character carrying the meaning " *to gather* " or " *assemble* " is still presently in use. The sign-etymology reveals the following development of morphological structures:

1.  , 2.  , 3.  , 4. 

1. A bird flies into the direction of the arrow on top of a tree. Numbers 2, 3 and 4 represent abbreviations within several thousand years of use. 4. expresses the present day used form in China and Japan.

§ 21 For each *RADICAL* the equivalent lemma or stem-word of at least ten National Language Systems has to be provided and implemented.

§ 22 Context sensitivity is limited down to a sequence law: Subject-Object-Predicate. Example: /father/son/moneytransfer/.

§ 23 *RADICALS* written in vertical strokes and *PARTICLES* written in brakes and vertical strokes are expressed by letter sequences without capital letters.

Example of rewriting conventional text into ALS-TEXT:
Text to be rewritten: *Newsweek*, February 5, 1973, page 1, top of the week, second column above: MACAO BOOMING OUTPOST, Page 23.:
Portugals pint-size colony of Macao is a geopolitical anachronism perched precariously on the flank of China. Newsweek correspondent Tony Clifton reports how Macao, once known as the world's wickedest city, is flourishing against all odds- and changing in the process. The preceding text is reduced to sequences of heteronoms or sentences according to ALS-Rules: Caution: This states an example of ALS-Rules and is not implemented for case study purposes!

1. /macao/outpost/boom
2. /macao/23/page
3. /portugal/macao/pint-size/colony/
4. /macao/geopolitical/anachronism/
5. /macao/china/flank/unsafe/sitting/

abstract, all ALS-Sentences or heteronoms exist of abstracts; with the exception of No. 7

6. /newsweek/correspondent/tonyclifton/macao/report/
7. /macao/ (pret.) /first/wicked/world/city/
8. /macao/ (against /sum/odd/flourish/
9. /macao/process/change/.

5.0. *A case study within a molecular scope of human knowledge*

5.1. *Archival Hard Storage by the ALS-System.*

The whole human knowledge is to compare with an immense Cluster of *RADICALS*. Each *RADICAL* represents one meaning. All of them are connected to each other following the graphic rules of a directed graph. The number of meanings is finite. Each day new meanings are added, because inventions and scientific detections are made. Thus the cluster grows daily. In order to obtain a 100% precise record of the Cluster, each *RADICAL* should be evaluated after its eight content- or subcriteria. These data of the criteria have to be implemented into a storage system. For reasons of rationality primarily facts should be stored and expressions with untrue contents omitted.

One molecular scope of the human treasure of knowledge is shown in Fig. 9. This *RADICAL-CLUSTER* in form of a directed graph explains the single connections between the *RADICALS*. If a Cluster in form of a cable-tree is given (P. KÜMMEL, 1972a, p. 65), compare also the loops in figures 4, 5 and 6 above, at each *RADICAL* the *DISTRIBUTION-* or *RADICAL-VALUE* is indicated by counting the number of cables in the loop.

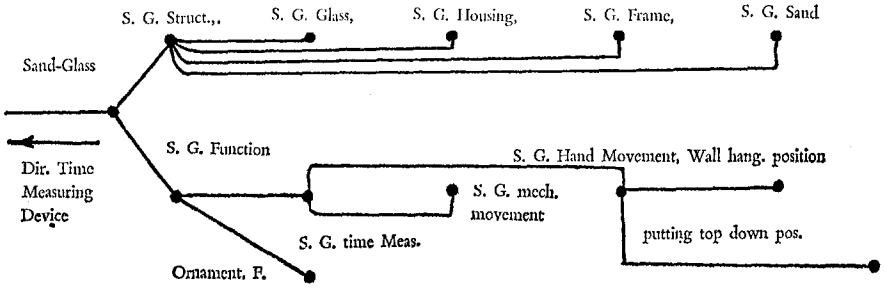


Fig. 9.

In Fig. 9 five concreta and eight abstracta are listed. The way they are implemented can be considered ripe for eternal implementation.

But for each *RADICAL* the enumerated sub-criteria must be given (compare 4.0!).

5.2. *ALS-fact retrieval in natural language.*

The procedure of Fact Retrieval is in this case not simulated with the help of a model but executed in realistic practice. The naturally grown graph in Fig. 9 as a "directed graph" allows questions and precise answers. The precision comes close to that demanded by patent office examiners. The operation of Fact Retrieval in the *RADICAL-CLUSTER* is compared to the function of a beetle, crawling up and down the branches of a bush, a cluster or a tree.

Possible questions and answers:
Thirteen meanings are cut off from the knowledge treasure and put into the system including their relations. The user might

1. ask about one meaning: He delivers his question in natural ALS-Text. For instance: /existence/(quest.)/sand-glass/function/. "/existence/ /(quest.)/" alone represents the ALS - equivalent for "*what is a ...?*", or, "*what about a ...?*". This constituent of a query "/existence/ /(quest)/" triggers immediately a first order retrieve circuit (Fig. 10), which is programmed to print all content criteria of one *RADICAL* in ALS-Text. The other constituent of "/sand-glass/function/" guides the first order retrieve circuit to its corresponding loop or *RADICAL* in the information graph. The content criteria of the heteronom /sand-glass/function/ are taken from the storage and are printed,

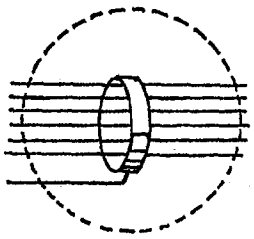


Fig. 10.
(first ord. *R. C.*)

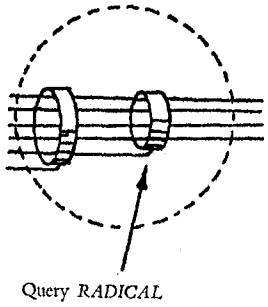


Fig. 11.
(second ord. *R. C.*)

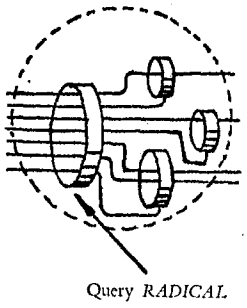


Fig. 12.
(third ord. *R. C.*)

2. ask, what does a meaning belong to? For instance in ALS-Text: /depend/(quest.)/sand-glass/function/. This triggers a " second order retrieve circuit " (Fig. 11). For the purpose of an answer all content criteria of the next upward loop or *RADICAL* (in this case " /sand-glass/ " are printed in ALS-Text,
3. ask, what does a concrete meaning consist of, respectively what subfunctions an abstract meaning can be divided in? This reads for instance in ALS-Text: /consist/(quest)/sand-glass/structure/. By " /consist/(quest)/ " or " /divisible/(quest/ " a third order retrieve circuit (Fig. 12) is triggered. All content criteria of the neighbouring ramificated *RADICALS* from the " Query - *RADICAL* " are printed in succession according to the decimal fracture of the *RADICAL*-Numbers.

For the procedures so far three questions different in function were possible:

	clear text:	ALS-Text:	implemented procedure:
1.	*what is a ...?*	/ existence / (quest) /	1st order retrieve circuit
2.	*to what belongs a ...?*	/ dependent / (quest) /	2nd order retrieve circuit
3 a.	*of what consists a?*	/ consist / (quest) /	3rd order retrieve circuit
3 b.	*into what functions divisible is a ...?*	/ divisible / (quest) /	3rd order retrieve circuit

If the preceding microscopic scope of human knowledge is extended, or further scopes including human relations are added, surely other questions must be available or made available as for instance: who?, by who?, etc. Also the *RADICALS* of many concreta must be subdivided and ramificated into concreta as well as abstracta.

5.3. *ALS-Automatic translation in natural language.*

In order to make use of the *INTERLOGGING-VALUE* of one *RADICAL*, the following jump has been made from the *RADICAL* /sand-glass/: A person associates with Sand-Glass an uncle in London, who, in former times, helped to construct a working model of a Sand-Glass. This uncle has a son studying in Paris who spends lots of money from his father. The interlogging string ends at the *RADICAL*: /father/ (Fig. 13).

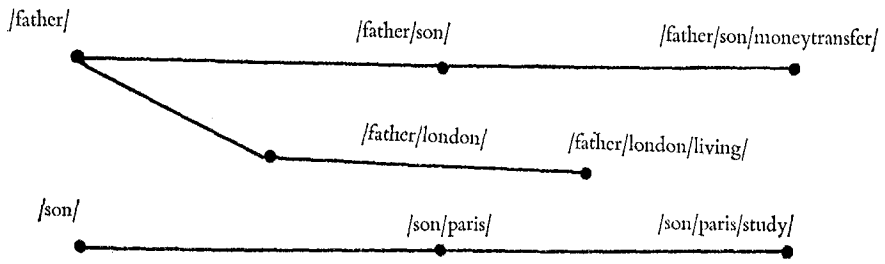


Fig. 13.

A second interlogging jump is made, when the sentence: *The father who lives in London, gives his son studying in Paris money.* Three ALS-Sentences are conceived:

1. /father/son/moneytransfer/
2. /father/London/living/
3. /son/Paris/study/

The procedure of Automatic Translation demands that each *RADICAL* to be translated must be implemented into the *RADICAL-CLUSTER* as in Fig. 13. All sub-criteria of a *RADICAL* have to be implemented by numbers too. Corresponding to the Fact Retrieval above in 5.2., each meaning or information unit implemented as a *RADICAL* into the system is additionally related to a set of numbers. These related identification numbers each point to a repertoire of nouns belonging to national languages. If, for instance, the system is furnished with ten different national languages, every single *RADICAL* handled must be numbered with a national identification number (P. KÜMMEL, 1972b, p. 486). From 1 to 10 for example the following national language systems might be put in sequence corresponding to their frequency of occurrence and importance by the user's per capita rate:

1. English	2. Chinese	3. Indian
4. Russian	5. Spanish	6. French
7. Portuguese	8. Japanese	9. German and
	10. Dutch	

Each meaning represented in the system by a *RADICAL* or loop with its eight content criteria, is international in character. Only the *RADICALS* can carry expressions of several national languages, not the *PARTICLES*. To obey this rule, words for question input can be

of ten different national origins. The system is programmed to mark each word with an additional national number, corresponding to the very nationality. Thus, ten different words link to one implemented meaning of one *RADICAL* and are connected to it.

The demand to translate into one of the ten above listed languages triggers a corresponding number between 1 and 10. Consequently ALS-Texts can be printed for answers in 10 different national languages. Because of internationally known syntactic patterns exclusively applicated in the ALS-Sequence of heteronoms, the nationality of each *RADICAL* in use does not affect the syntactical momentum involved. This leads to a simplification of the whole act of *AUTOMATIC TRANSLATION*. The entire procedure is solved by exclusively translating the *RADICAL*-Expressions for the whole information in form of heteronoms. If needed, the number of different national languages implemented into the system can be extended voluntarily above ten.

Automatic Translation of different national languages according to the ALS-System is only possible if all information units (meanings or *RADICALS*) concerned are implemented in the very patterns of the Automatic Fact Retrieval under § 2 and after the ALS-Rules. All texts to be handled must follow stereotype syntactical features obeying strict eliminations of syntagms with low frequencies of occurrence as described in the ALGORITHM OF LIMITED SYNTAX (ALS).

REFERENCES

Y. Bar Hillel, *Communication and Argumentation in Natural Language*, Milano 1970.

S. Beer, *Kybernetik und Management*, Hamburg 1963.

L. Bloom, *Language Development*, Cambridge (Mass.) 1970.

L. Bloomfield, *Language*, London 1935, 566 p.

R. Brown, with coauthors, *Psycholinguistics, Three processes in the Child's Acquisition of Syntax*, New York 1970.

N. Chomsky, *Aspects of the Theory of Syntax*, Cambridge (Mass.) 1965.

J. H. Greenberg, *Universals of Language*, Cambridge (Mass.) 1968, pp. 73-113.

R. v. L. Hartley, « Bell Systems Technical Journal », VII (1928).

P. Kümmel, *Die Sechsschreibung der chinesischen Schriftzeichen als Einführung in den Fragenkreis des willkürlichen Ausdrucks*, in « Grundlagenstudien aus Kybernetik und Geisteswissenschaft », Quickborn bei Hamburg, VII (1966) 4, pp. 109-118.

P. Kümmel, *Ideographie - Funktion des sichtbaren Ausdrucks*, in « Grundlagenstudien aus Kybernetik und Geisteswissenschaft », Quickborn bei Hamburg, VIII (1967 a) 2, pp. 63-66.

P. Kümmel, *Ideography - Function of Voluntary Expression*, in *Transactions of the International Conference of Orientalists in Japan*, No. XII, The Institute of Eastern Culture, Tokio 1967 (b), pp. 36-50.

P. Kümmel, *Hanzeichen (Chinese Characters) ohne ideographische Funktion*, in « Grundlagenstudien aus Kybernetik und Geisteswissenschaft », Quickborn bei Hamburg, IX (1968 a) 2, pp. 48-57.

P. Kümmel, *Betrachtungen zur Ausdrucksanalyse*, in « Zeitschrift für Sprachwissenschaft Phonetik und Kommunikationswissenschaft », Berlin, XXI (1968 b) 3-4, pp. 232-244.

P. Kümmel, *Struktur und Funktion sichtbarer Zeichen*, Quickborn bei Hamburg, 1969, 233 p.

P. Kümmel, *Wertbestimmung von Information, zur qualitativen Analyse des Inhalts natürlicher Sprachen, Grundlagen der Nichtnumerik*, Tübingen 1972 (a), 90 p.

P. Kümmel, *Formalisieren von Inhalt natürlicher Sprachen durch Deutervalenzen und Deuterkriterien*, Lecture Notes in Economics and Mathematical Systems, Berlin, Heidelberg, New York, (Vortrag bei der GI-Tagung 1972 in Karlsruhe), 1972 (b), pp. 480-489.

Mc. D. Neill, *The Capacity for the Ontogenesis of Grammar*, in D. L. Slobin (ed.), *The Ontogenesis of Grammar*, A Theoretical Symposium, London, New York 1971.

M. L. Minsky, *Semantic Information Processing*, Cambridge (Mass.) 1968, pp. 1-32.

G. Salton, *Automatic Information Organization and Retrieval*, London, New York 1968.

H. Schnelle, *Language Communication with Children, Toward a Theory of Language Use*, in Y. Bar-Hillil (ed.),

Pragmatic of Natural Languages, Dordrecht (Ne.), 1971, pp. 173-193.

C. E. SHANNON, *Recent Contributions to the Mathematical Theory of Communication*, Urbana (Illinois) 1964.

W. WEAVER, *Recent Contributions to the Mathematical Theory of Communication*, Urbana (Illinois) 1964.

T. WINOGRAD, *Understanding Natural Language*, London, New York 1972.