

CONTENT GUIDED ANSWER SEARCH SYSTEM
FOR NATURAL LANGUAGES

Peter Kümmel
Institut für Informatik
Stuttgart University, FRG

SUMMARY:

Portions of natural language content must be directly related to memory functions in computers. A system should be selected which does not only fit theoretically but also can be practically⁶ applied, due to its "natural" character. It represents a difficult task to cut away a usable model thesaurus from the huge meaning treasure of natural language. Air Traffic Control Language, restricted to Ground Control, serves this purpose. Due to severe realtime requirements⁴ a limitation to conventional computers has been avoided. Exclusively^{11,12,13} content analogous memory structures^{3,10} guarantee an automatic answer search, which obeys the realtime obligations of 4-6 syllables/s speech velocity and 0.5-1.5 s response interval.

GLOSSARY:

Agglutination, agglutinated morphology = agglutinated expression structures of language = expression mixed with syntax particles, answer actualization, answer pool (AP), answer search logic (ASL), content unit = meaning, deuter = SVA which is defined after six sub-criteria, deuter-matcher, deuter-criterion = one of the six sub-criteria of one deuter, deuter string synthesizer DSS, feature extraction, flexion, content-unit = meaning = deuter plus syntagms, isolated morphology = language expression existing mainly of deuter sequences.

key feature of content KFC, KFC-matcher, lemma/ta = word stem, morphology = expression structure of language, progressive agglutination, suffix = pre- or postponed syllable to word stems, SVA = substantive, verb, or adjective, if possible in lemma shape, syntagm = consisting of syntax particles and syntax rules, syntax particle, syntax rule, twin KFC = a double KFC implemented into a ROM to match potential "input-KFCs".

CR Categories: 3.65, 3.71, 3.74, 3.75, 3.81, 5.32, 5.6, 7.3.

1 INTRODUCTION:

To substitute one partner in natural language dialogs by a computer, a model dialog should be chosen, which uses besides user utterances also system alerts for input. These alerts are triggered by the user's action. The introduced "Standard 80" thesaurus meets these demands. It concerns ground and ramp control activities in ATCL dialogs between pilots and air controllers. In those Air Traffic Control Language dialogs taxy procedures by the user, in this case the pilot, are in addition monitored by sensors at the apron. Thus, besides "primary inputs" by pilot utterances also "secondary inputs" by the system help the content guided answer search mechanism. The ATCL at international airports has been reduced to one

particular language - English - already in the fifties. In a minimum of time elapse a maximum of information flux between the dialog partners is needed. This postulate causes all grammatical components with a low level efficiency rate to be abolished naturally.

Considering most recent structural evaluations, formalization results of natural language⁶ led to the determination of: "key features of content; KFCs.^{8,9} By a matching procedure of input KFCs with twin KFCs stored in a ROM, prepared answers can be triggered. But, a match instruction can be also used to trigger a sequence of deuter, which are implemented in a time-analogous read-only array. The trigger procedure of deuter sequences is faster than that of prepared answers. Especially if more voluminous thesauri are used.^{1,2} But, up to the present, an actualizing of deuter sequences is still missing proper morphological agglutinations. Automatically triggered deuter sequences appear in isolated morphologies. For this purpose an "algorithm of progressive agglutination"⁶ is under consideration, by which syntax particles and additional syntax rules can be mingled into deuter sequences. For the ATCL however deuter sequences without too many syntagms are quite sufficient.

2 "STANDARD 80" A MODEL THESAURUS

2.1 Alphabetical Sequence

100 meanings = 82 deuter plus 18 syntax particles, ATCL, Ground Control, 121.9 Mc, Stuttgart International Airport: in alphabetical order, syntax particles underlined:

affirmative, airborne frequency, altimeter setting, approved, apron, 6 Armycopter, backtrack, behind, 9 brake malfunction,

calling, call sign, caution, center line, Cessna 172, change to 118.8, charlie, circle/circling, clearance, cleared to, construction work, contact radar 125.05, contact tower 118.8, continue, correct, 25 crossing, D-EIPV, departing, departure route 28, exit point, expect level change en route, 31 expedite, fire truck, flight level, follow, follome, 36 for, go ahead, 38 goodbye, Hamburg airport, helicopter, 41 hover, information, 43 intersection, level change en route, Lima, local flight, Lufthansa Boing, 48 Lufthansa 792, maintain, military aircraft 50, next, 52 November departure, on grass, 125.05, opposite direction, out, outbound, 58 over, Pan Am Clipper, parking area, permission, please, position, 64 present, QNH 1019 mb, 66 QNH 1009 mb, radar 125.05, ramp, ready to, readback, refueling, request, roger, runway 26, 75 runway 08, say again, slow down, special VFR, standby, start up clearance, start up engines, stay clear off, stop, 84 Stuttgart Tower, takeoff, taxi clearance, taxi/taxying, this is, to the West, 90 type of aircraft, 91 via departure route, wait, we are, what, when airborne call radar 125.05, when ready, wilco, 98 wind, 99 you are, 100 0493 PA.

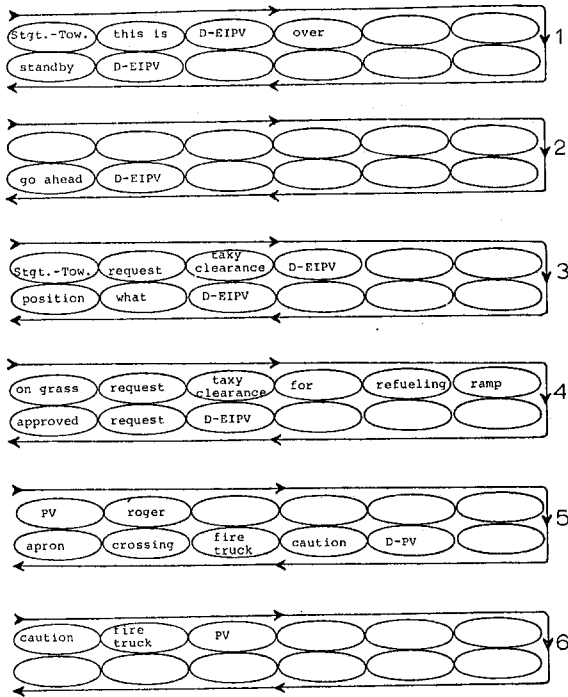
Appearing numbers within the deuter accumulation in alphabetical order, are interim counts from 1 to 100.

18 Syntax Particles:

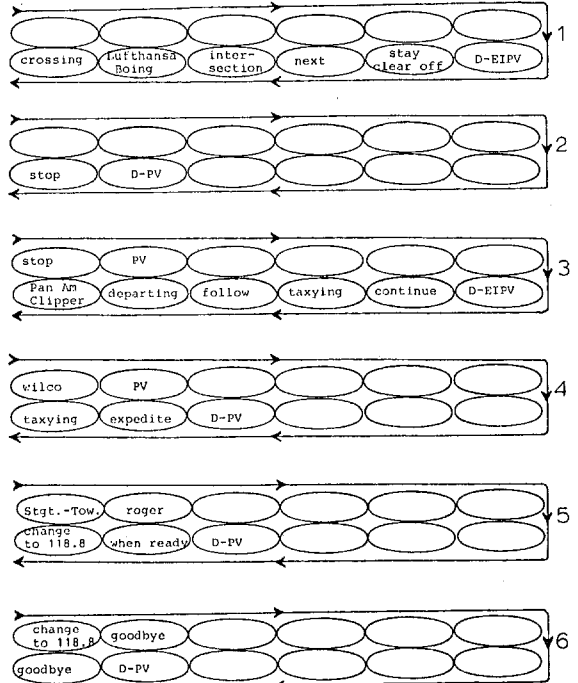
1. affirmative, 2. approved, 3. correct,
4. for, 5. go ahead, 6. goodbye, 7. next,
8. out, 9. over, 10. please, 11. roger,
12. standby, 13. this is, 14. we are, 15. what, 16. when ready, 17. wilco, 18. you are.

2.2 The Eight Partial Dialogs:

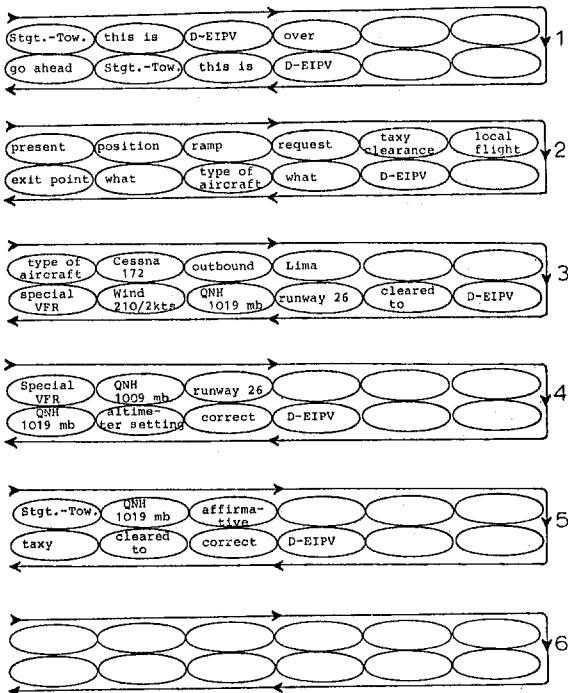
Partial Dialog No. 1: VFR Clearance: Grass - Ramp
 Ground Control, 121.9 Mc, Stuttgart (Stgt.) International Airport
 Aircraft: Cessna 172, D-EIPV



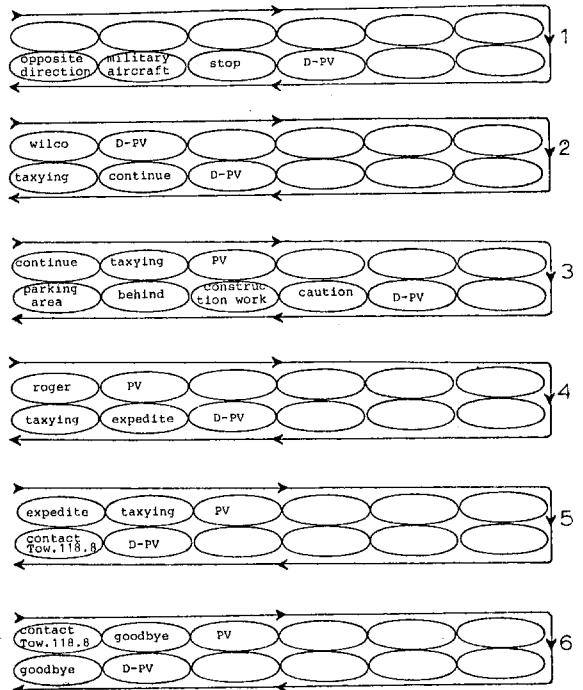
Partial Dialog No. 3: VFR Taxi Procedure: Ramp - Runway I
 Ground Control, 121.9 Mc, Stuttgart (Stgt.) International Airport
 Aircraft: Cessna 172, D-EIPV



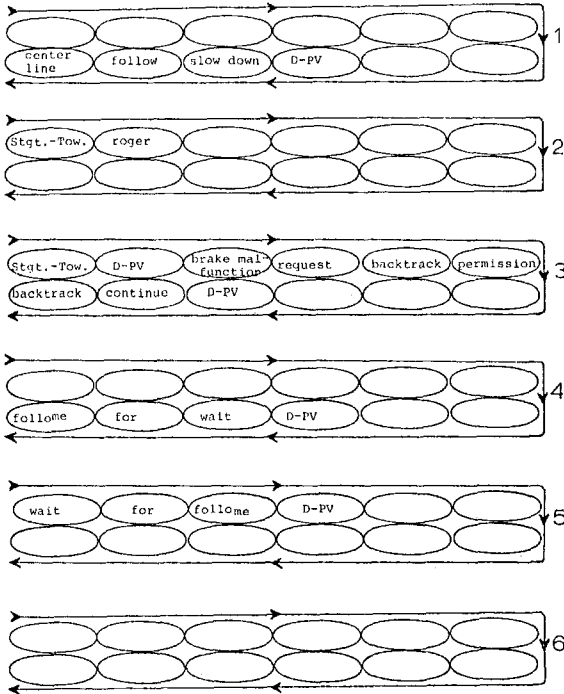
Partial Dialog No. 2: VFR Clearance: Ramp - Runway
 Ground Control, 121.9 Mc, Stuttgart (Stgt.) International Airport
 Aircraft: Cessna 172, D-EIPV



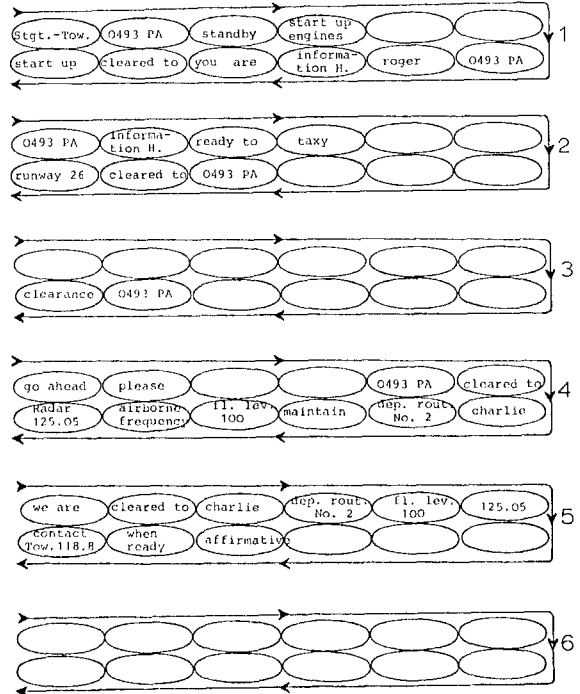
Partial Dialog No. 4: VFR Taxi Procedure: Ramp - Runway II
 Ground Control, 121.9 Mc, Stuttgart (Stgt.) International Airport
 Aircraft: Cessna 172, D-EIPV



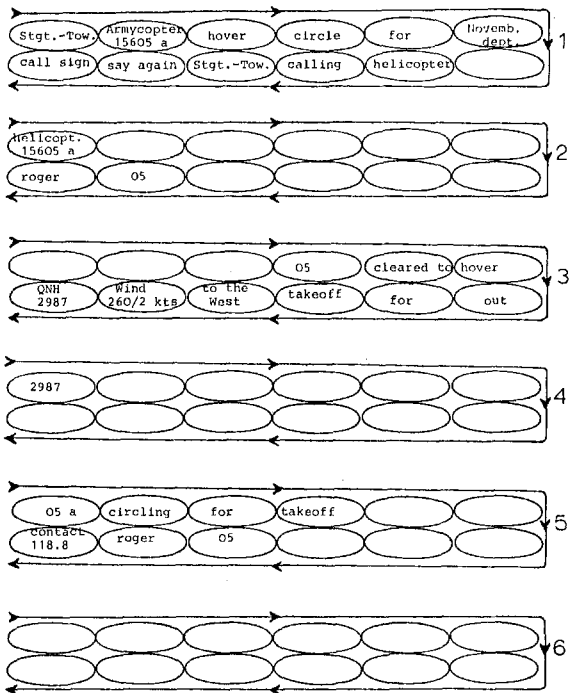
Partial Dialog No. 5 : VFR Taxi Procedure: Ramp - Runway III
 Ground Control, 121.9 Mc, Stuttgart (Stgt.) International Airport
 (Brake Malfunction). Aircraft: Cessna 172, D-BPV



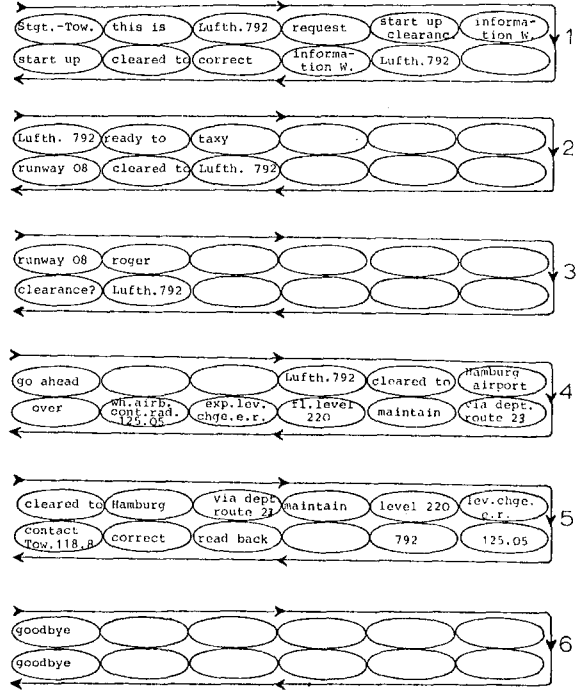
Partial Dialog No. 7 : IFR Taxi- and En Route Clearance: Ramp-Runway
 Ground Control, 121.9 Mc, Stuttgart (Stgt.), International Airport
 US Army Flight to Frankfurt



Partial Dialog No. 6 : VFR Hover Procedure to Helicopter Circle and
 Takeoff, Ground Control, 121.9 Mc, Stuttgart, International Airport
 US Army Helicopter: UH 1, 15605 a



Partial Dialog No. 8 : IFR Taxi- and En Route Clearance: Ramp-Runway
 Ground Control, 121.9 Mc, Stuttgart (Stgt.), International Airport
 Lufthansa Domestic Flight: Stuttgart-Hamburg



2.3 Major Syntax Rules in the Standard80

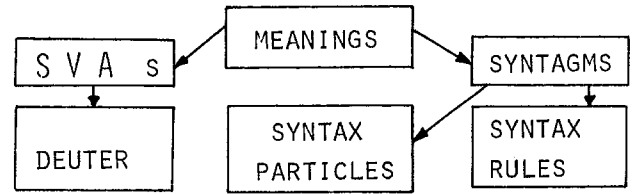
- 1) Talk and requests are only permitted, if the line is empty.
- 2) Confirmation of instruction is possible by repetition of the instruction, or major parts in it.
- 3) Subject/Predicate-Sequence (SP).
- 4) Confirmation of partner utterance by stating the last two letters of the aircraft identification.
- 5) Abbreviation of high frequency phrases to short expressions: examples: "will comply with" to "wilco" and "understood" to "roger".
- 6) Extension of easy misunderstood high frequency syntax particles, e.g. "no" to "negative" and "yes" to "positive" or "affirmative".
- 7) All dialogs should not contain more than six cycles (dialog cycles).
- 8) All dialogs should not contain more than six input- and six output deuter except for en route clearances.
- 9) Absolute obligation of pilot to confirm QNH-data.
- 10) Urgent obligation to confirm the: "change to 118.8" or "118.8" the runway frequency as a sign, the dialog is finished.
- 11) After the second cycle the aircraft identification should be abbreviated from e.g "D-EIPV" to "D-PV".
- 12) Each tower utterance should possibly commence with the aircraft identification.

3 LANGUAGE COMPONENTS AND THEIR TERMINOLOGY

3.1 Content Units (meanings)

A natural language L_n expression thesaurus can be split into a) a group of substantiva, verba, and adjectiva,⁶ the SVAs and b) a group of syntagms (Fig. 1). If the SVAs, possibly in lemma shape, are

FIG. 1
ANALYSIS OF " MEANINGS "



defined after their sub-criteria, the deuter-criteria,^{6,7} they are called "deuter".

3.2 The Six Deuter-Criteria:⁶

- 1) Identity: given by DC-enumeration.
- 2) Age: absolute and relative, also in comparison to dialog cycles.
- 3) Association: A_v and A_h , vertical and horizontal, A_v is to compare with edges in a directed graph.
- 4) Frequency: absolute and relative, is in smaller thesauri related to the significance value.
- 5) Significance: has influence on syntax rules and their deuter, e.g. No. 9, (2.3) very high value.
- 6) Truth: also existent in smaller dialogs. If untrue utterances by the pilot are given as e.g. "mercedes benz" after the tower inquiry: "type of aircraft?" or, "profane" utterances like: idiot, nonsense, etc. They are not permitted and thus "untrue".

3.3 Syntagms

3.3.1 Definition: Syntax Particle

Excluding SVA-Lemmata, all suffixes, flexions, and remaining words of an L_n thesaurus are considered Syntax Particles.

3.3.2 Definition: Syntax Rule

They include rules of word sequences: S/P, S/O/P etc. and can be directly related to rules of dialog functions.

4 STRUCTURE AND FUNCTION OF DIALOG CYCLES

4.1 User inputs: a) identific., b) identific.+request, c) request, d) instruction confirmation, e) wrong instruction confirmation and f) unclear identific.

4.2 System inputs: a) interrupt, b), c), d) three different "wrong user actions".

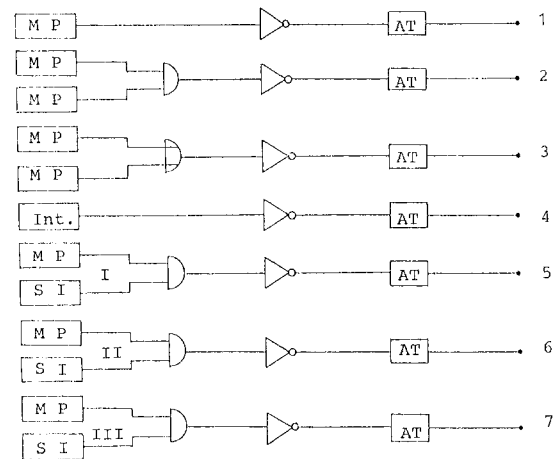
4.3 System Outputs: a) instruction to continue (4.1a), b) instruction + permission (4.1b,c,d), c) correction (4.1e), d) request for correct identific. (4.1f), e) standby instruct. (4.2a), f) correction to (4.2b), g) correction to (4.2c), and h) correction to (4.2d).

4.4 Determining the KFCs:
Depending from the deuter-criterion: age-value, the dialog cycles 1-6 from old to new are brought into relation to the sum of all potential partial dialogs, in this case 8. Among all cycles of the same age, frequency values and redundancies of single deuter are checked. Only a few deuter among all input data have the "switching ability" to select one of the 8 potential next cycles. These account for less than 20% of all input data. They are called "key features of content", KFCs.

5. ANSWER SEARCH LOGIC FOR PREPARED ANSWERS (ASL)

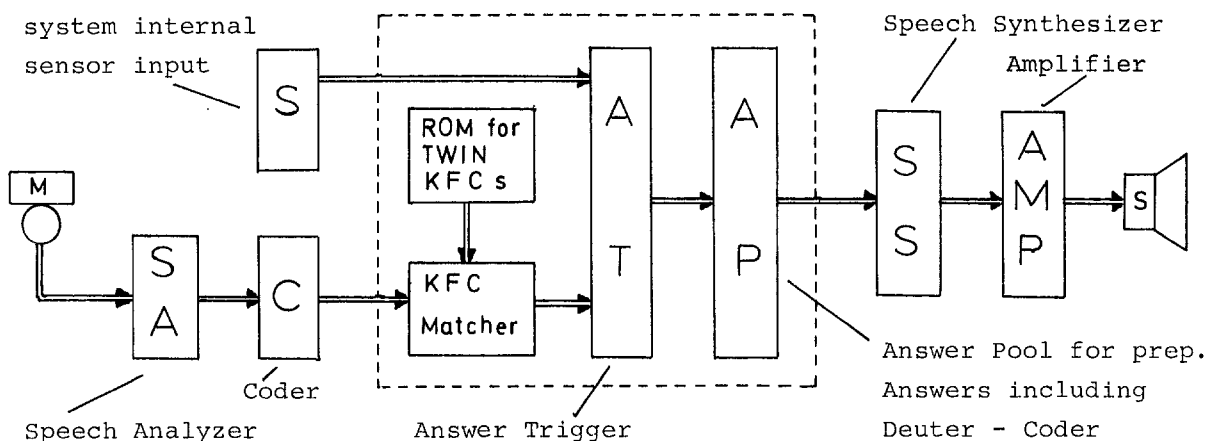
5.1 Match Procedures MPs of KFCs:
According to 4.4, all potential KFCs are implemented as "twin-KFCs" in a twin-ROM. The implementation array is time-analogous. As the input-KFCs appear also after the deuter-criterion age-value, a realtime MP will be quite simple.

FIG. 2 SEQUENTIAL NETW. OF AT



5.2 Trigger of prepared answers AT
As each dialog cycle of the eight dialogs consists of a) inputs and b) outputs, the dialog outputs (air controller utterances) can be easily copied for pre-

FIG. 3 ASL - ARCHITECTURE :



pared answers. They are triggered, if KFC match procedures MPs and system internal sensor inputs passed a sequential network AT. According to the 8 partial dialogs PDs, 7 switching combinations are possible: 1) MP of 1 input KFC (PD6), 2) MP of 2 input KFCs (PDs 1,2,7,8), 3) MP of one among 2 input KFCs (e.g. in cycles 2-6), 4) interrupt (PD 1), 5) MP of 1 input-KFC + 1 of 3 systeminternal alerts No I, 6) like 5) No II, 7) like 5) No III.

5.3 Sequential Network of AT

According to 5.2, for each Partial Dialog 7 potential switching combinations are possible (Fig. 2).

5.4 Components of the ASL

The main units consist of a) a ROM for twin-KFCs, b) a KFC-matcher as a comparator device to match input-KFCs with twin-KFCs, and c) an AT (Fig. 3).

6 DEUTER STRING SYNTHESIZER DSS

6.1 Function of a DSS

According to function- and memory redundancies in ASLs, (Fig. 3), the units:

- 1) KFC Twin ROM,
- 2) KFC Matcher
- 3) Answer Trigger, and
- 4) Answer Pool for Prepared Answers,

are combined in a "DSS-Logic". Not only twin-KFCs but all potential deuter in all dialogs are stored in the DSS-Logic.

This, in a way, substitutes the Answer Pool for prepared answers. The read only memory implementation must be in geometric "time logic". Thus an electrical activation of one matched twin-KFC triggers a sequence of surrounding stored deuter. This triggering is executed by a "switch thru-", a "put through connection" procedure. Each actualized deuter in such a "switch thru sequence" is si-

multaneously transformed into an audio output, a synchronous "read out". The "switch thru association" becomes "loud thinking". This fits a logic answer to the input-KFCs. Major functions of the DSS are:

- 1) All twin-KFCs serve potential match procedures with input-KFCs.
- 2) After an input-KFC is matched, the matched twin-KFC is activated as "the first deuter" in a triggered deuter sequence.
- 3) The length of the sequence will be limited from 3 to 10 deuter.
- 4) Each connected deuter in a triggered deuter sequence is allocated with 2-3 seconds time, in which the adequate English utterance can be expressed by a speech synthesizer device.
- 5) The "switch thru connection" to neighbouring deuter in the deuter sequence continues only, after the English word of the last deuter is spoken out.

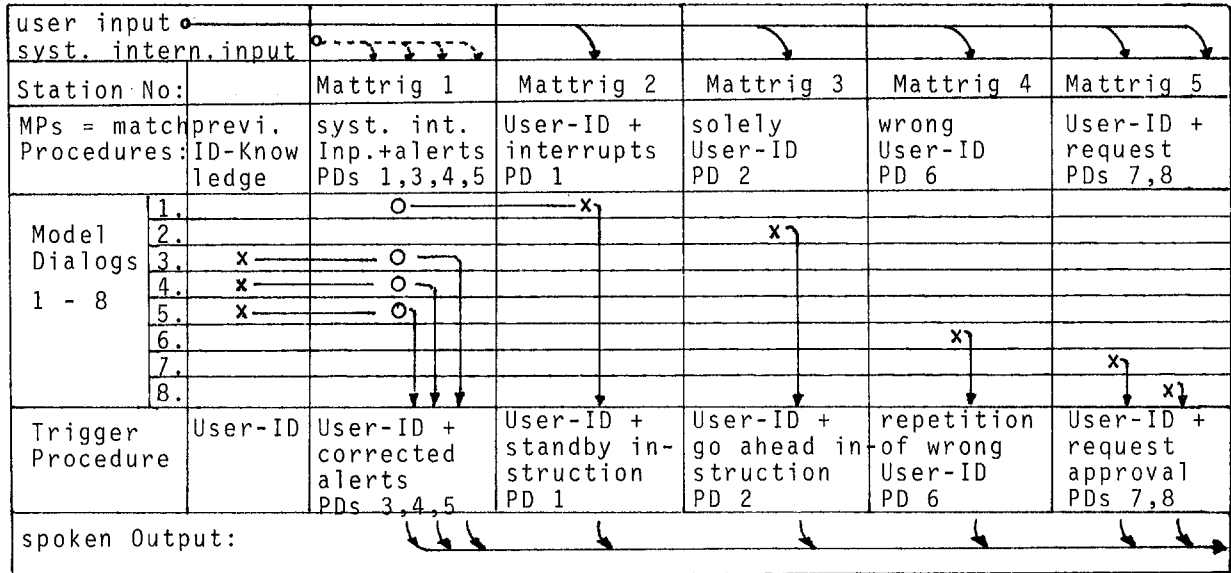
6.2 Architecture of a DSS

The array of DSS content is divided into dialog cycle sections. For each of them five or more KFC-Match/Trigger Stations (Matrigs) are provided. Structure and organization of function units in all first dialog cycles of the eight model dialogs are shown in Fig. 4: If such sections for the remaining dialog cycles 2-6 of all potential dialogs are compiled in an analogous way and put together, a further elimination of redundant portions will help to create a complex DSS architecture.

Acknowledgements:

By courtesy of the "Bundesanstalt für Flugsicherung" in Frankfurt/M, the author has been permitted to copy tapes at the air traffic control center at Stuttgart

FIG. 4 SECTION ONE OF ALL FIRST DIALOG CYCLES



Echterdingen, International Airport for statistical thesaurus evaluations⁵. A research project sponsored by the "Deutsche Forschungsgemeinschaft" in Bonn, under the topic: "Answer Searcher" is supervised by the author.

7 Conclusion:

Function simulation of one partner in nat. lang. dialogs by a machine demands the following major procedures:

- 1) KFC-matching of Input-KFCs with Twin-KFCs implemented in a ROM,
- 2) Answer Trigger according to KFC match procedures in devices with prepared answers ASLs, and
- 3) Actualization of Deuter String Outputs in DSSs after KFC matchings.

More than expected, complex evaluations of nat. lang. phenomena were necessary for memory array purposes. Among them particularly the six deuter-criteria. But also a sufficient familiarity with the applied thesaurus was necessary. Since 16 years the author holds a valid PPI license and executes active flying.

References:

- 1.ERMANN,L., Org. of the HEARSAY II Speech Underst. S., Work.Pap.,CMU 1974
- 2.FENNEL,R., co-author, see No. 1, 1974
- 3.HAYES-ROTH,F., Focus of Attention in a Distr. Log. Speech Und. Syst.,CMU 1976
- 4.JELINEK,F., Continuous Speech Recogni. by Stat. Meth., IEEE,Vol.64 ,No 4,1976
5. KÜMMEL,P., Hybrid Memory f. Autom.Air Traffic Control, IMACS, North Hol.1978
- 6.KÜMMEL,P.,Formalization of Natural Languages, Springer, Berlin-New York,1979
- 7.KÜMMEL,P.,Information Analys. by Formalizing Content of Nat. Lang.,in: Meth. of Oper. Research, IIIrd Symp., 1979
- 8.KÜMMEL,P., Answer Search Interface, Proc. of CompCon, Fall 79, Wash., 1979
- 9.KÜMMEL,P.,Zur Implementierung v. Dialoginhalten f.e. Relat. Abfragesprache Informatik Fachber. Springer, 30, 1980
- 10.LESSER,V., co-author, see No 3, 1976
- 11.LEILICH,H., The Search Processor, Inform. Fachber. TU Braunsch.7801, 1978
- 12.STIEGE,H., co-author, see No 11, 1978
13. ZEIDLICH,H., co-author, see No11,1978