Situácia sa značne mení, ak sa ako vstup použije spontánny rečový prejav, ktorý je do textovej formy prevedený počítačom. Ku vstupu sa pridávajú ďalšie chyby, ktoré sú zapríčinené zlým rozpoznaním reči.

Medzi najznámejšie z množstva problémov, ktoré vznikajú pri analýze spontánnej reči pomocou počítača, patria napr.:

- neznáme alebo zle rozpoznané slová,
- vyplnené pauzy (hm, ah, ...),
- reštarty – opakovanie slov, alebo fráz,
- citoslovcia,
- elipsy,
- negramatické konštrukcie.

Prvý z uvedených problémov je problém rozpoznávača, ostatné sa však riešia práve sémantickou analýzou. Správne riešenie uvedených javov je problém pre väčšinu tradičných analyzátorov.

Existujú dva principiálne prístupy na riešenie tohto problému: gramatika analyzátora je rozšírená o modelovanie vyššie uvedených javov alebo lingvistická analýza sa navrhuje s ohľadom na danú problematiku (Eckert – Nieman, 1994).

V našom prípade bol zvolený druhý prístup, založený na analýze kľúčových slov a ich kombináciách (Convington, 1994). Výsledkom je sémantická reprezentácia vstupnej vety. Pokiaľ nie je možné jednotlivé analyzované časti vety pokryť žiadnou gramatikou, čo znamená, že vstupná veta nie je gramaticky správna, výstupom sú parciálne výsledky.

Identifikácia kľúčových slov

Pri analýze kľúčových slov sa neberie veta ako celok, ale každé slovo sa spracúva samostatne. Vstupné slovo sa buď akceptuje a priradí sa mu jeden z 11 možných sémantických typov, alebo sa považuje za „jazykový šum" a vynechá sa. Pre použitie vo vlakovom informačnom systéme rozoznávame nasledujúce typy:

- východisková stanica,
- cieľová stanica,
- relatívny čas (*večer, ráno, poobede* ..),
- absolútny čas (*o tretej, 4.20* ..),
- relatívny dátum (*o týždeň, v pondelok* ..),
- absolútny dátum (*3. januára, 21 6* ..),
- predložky (*pred, okolo* ..),
- podmieňovacie spojky (*ak, keď*),

- súhlas (*áno, správne* ...),
- nesúhlas (*nie, nesprávne* ..),
- pozdrav (*dopočutia, čaf* ..).

Vstupná veta je reprezentovaná zoznamom, kde každý prvok predstavuje jedno slovo vstupnej vety. Zoznam je spracovaný sekvenčne. Prvý prvok zoznamu sa testuje na príslušnosť k niektorej skupine *kľúčových slov* Testy sa delia do štyroch základných skupín:

- testy na cieľovú a východiskovú stanicu,
- testy na časové údaje,
- testy na dátumové údaje,
- ostatné testy (predložky, spojky ...).

Ak testovaný prvok patrí do niektorej zo skupín, pridá sa s priradeným typom k výstupnému zoznamu. Ak je prvok neznámy, vynechá sa a analýza pokračuje ďalej až do spracovania posledného prvku zoznamu. V niektorých prípadoch je existencia prvého prvku zoznamu viazaná na existenciu ďalšieho, ktorý sa nachádza v ešte nespracovanej časti. V takom prípade sa spracuje zvyšok zoznamu. Ak sa v ňom takýto prvok vyskytuje, vymaže sa a vytvorí s prvým prvkom jeden z typov. Tieto prípady sa vyskytujú hlavne v určovaní východiskovej a cieľovej stanice.

Testy na cieľovú a východiskovú stanicu

V testoch na cieľovú a východiskovú stanicu sa vo väčšine prípadov predpokladá spojenie názvu stanice s predložkou 'z' alebo 'do'. V takomto prípade sa dá jednoznačne určiť cieľová, resp. východisková stanica. Problém nastáva, ak je názov stanice v spojení s predložkou 'v':

1. *Potreboval by som byť dnes večer o piatej v Prešove.*
2. *Som v Prešove*

V týchto prípadoch nie je možné len na základe predložky stanicu jednoznačne zaradiť. V prvom prípade ide o východiskovú stanicu, v druhom o cieľovú. Na jednoznačné určenie sa používajú ďalšie slovné druhy. Východisková stanica sa deteguje pomocou tvarov sponového slovesa, avšak až potom, ako sa vylúči možná zámena za stanicu cieľovú. Z tohto dôvodu sa pri vyhľadávaní cieľovej stanice používajú rozličné tvary sponového slovesa. V súvislosti s východiskovou stanicou je to zvyčajne prítomný čas oznamovacieho spôsobu. Pri cieľovej stanici sa používa buď neurčitok '**byť**'s modálnym slovesom (*chcem byť, mám byť*), alebo podmieňovací spôsob (*chcel by som byť*).