

někdy tzv. pořad „strukturní“.⁹⁾ Chce se tím naznačit, že vedle „vnější“, lineární organizace složek textu existuje ještě jiná jejich organizace, skrytější, avšak závažnější. Podobně jako v pojmu lineárního pořadu jsou zahrnuti dva logicky nezávislé principy (antepozice — postpozice; postavení kontaktní — distantní),¹⁰⁾ lze aspoň dva takové principy rozoznat i v pojmu pořadu strukturního.

Vezmeme např. českou větu *Otec dočetl dlouhý dopis*. Může tu jít jednak o to, že jsou některé dvojice sl. tvarů spjaty vztahem synt. závislosti, jednak o to, že k sobě sl. tvary různé patří v tom smyslu, že se sdružují do jistých celků, tyto celky do vyšších celků atd.

2.1. Přistoupí-li v prvním případě omezení, že každý sl. tv.,¹¹⁾ popř. s výjimkou jednoho, nezávislého,¹²⁾ musí záviset právě na jediném sl. tvaru jiném, budeme hovořit o *k o n c e p c i* A. Z hlediska této koncepce si můžeme syntaktickou stavbu naší věty znázornit graficky takto:¹³⁾



Je jasné, že koncepce A zahrnuje zhruba společné rysy tradiční již u nás koncepcie syntaktických (tzv. větných) dvojic a koncepce slovních spojení (rus. slovo-socetanie).¹⁴⁾

2.2. V druhém případě budeme mluvit o *k o n c e p c i* B. Ta odpovídá zhruba tzv. analýze podle bezprostředních složek (N. A. Slijusareva, l. c.). Nejčastěji je ta podoba koncepce B, v níž se uplatňuje princip kontaktního postavení uvnitř zmíněných celků a princip dvojčlennosti těchto celků. Dále budeme koncepci B rozumět především tuto její podobu. Poznamenejme, že zatímco princip dvojčlennosti je tedy v koncepci B samostatnou složkou, je v koncepci A princip dvojčlennosti vztahu synt. závislosti logicky spjat se základní myšlenkou této koncepce. Z hlediska koncepce B si můžeme syntaktickou stavbu naší věty znázornit graficky např. způsobem u nás dost obvyklým:



⁹⁾ Viz např. L. Bloomfield, *Language*, London 1955, 210; L. Tesnière, *Éléments de syntaxe structurale*, Paris 1959, 19—20.

¹⁰⁾ S pomocí těchto pojmů se dají definovat pojmy: postavení začátkové, koncové a středové.

¹¹⁾ Srov. I. I. Revzin, *Formální a sémantičtější analýza syntaktických svazek v jazyce*, sb. Primenění logiky v науке i technice, Moskva, 1960, 130—131 (srov. i rec. K. Berký O některých aplikacích moderní formální logiky v jazykovědě, SaS 22, 1961, 199).

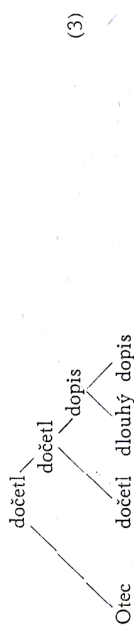
¹²⁾ Srov. L. Tesnière, op. cit., 14.

¹³⁾ Niž položený sl. tv. ze dvou tvarů spojených úsečkou je závislým členem dvojice, výše položený jejím členem řídicím. (Z grafu lze rekonstruovat slovosled analyzované věty). Viz u nás V. Šmilauer, *Novočeská škálda*, Praha 1947, 415n.

¹⁴⁾ Obě tyto poslední koncepce se od sebe liší především terminologií a jejich použitím je mnohdy spojeno s jistými uzančeními při postupu výkladu. Pro koncepci syntaktických (větných) dvojic je příznačné rozlišování různých typů synt. závislosti, srov. J. Bauer, *Několik poznámek o pojmech slovního spojení, větná dvojice a syntagma*, Sb. prací filos. fak. brn. univ. r. 1, č. 1—2, ř. jaz. (A), č. 1, zejm. 41—43. — Podle výše uvedeného omezení není v koncepci A např. protějšek vět-ných členů s dvojitou závislostí, srov. S. Ja. Fitalov, *O modelování syntaktické struktury*, Dokl. 61, vyp. 10, s. 10.

Jde tu např. o to, že *dlouhý* a *dopis* vytvářejí celek *dlouhý dopis* a ten spolu s *dočetl* celek *dočetl dlouhý dopis*. Koncepce B je matematicky prostudována mnohem více než koncepce A (viz pozn. 32 a 33).

2.3 Uvedme ještě graf sl. tv. *reprezentace* (rus. *представление*, angl. *representation*) této věty podle obou koncepcí.¹⁵⁾ Je na něm vidět, jak se obě koncepce doplňují.¹⁶⁾



3. K o n c e p c e A. Z hlediska této koncepce jde při zjišťování syntaktických vztahů přirozeně o to, zjistit pro každý závislý sl. tv. příslušný sl. tv. řídicí, tedy jinými slovy, sestrojít graf stejného typu jako je graf (1), srov. L. Tesnière, op. cit. v pozn. 9.

Úkony, které vedou k takovému výsledku, musí být, jak jsme se již zmínili, formální. Není ovšem možno sestrojít formální postup synt. analýzy pro všechny jednotlivé sl. tvary, nýbrž jen pro jisté jejich množiny, pro jisté synt. charakteristiky. Konkrétním sl. tvarům musí být tedy jednoznačně přiřazeny jisté takové synt. charakteristiky. Synt. analýza se pak provádí na posloupnosti symbolů synt. charakteristiké vzniklé z analyzované věty (posloupnosti sl. tvarů) záměnou každého sl. tvaru symbolem příslušné synt. charakteristiky.¹⁷⁾ Posloupnosti těchto symbolů budeme nazývat S-strukturami.¹⁸⁾

3.1 Nutným předpokladem pro sestavení algoritmu synt. analýzy, který ze S-struktury věty zjistí synt. závislosti mezi sl. tvary, je a) seznam jistých synt. tříd sl. tvarů,¹⁹⁾ b) seznam elementárních synt. konstrukcí, tzv. konfigurací (vyjádřených v terminologii těchto tříd),²⁰⁾ c) jisté poznatky slovosledné. Jak uvídíme (srov. 3.5.), lze a) a b) jistým způsobem upravovat (touto formulací nechceme ovšem nic tvrdit o konkrétním postupu prací při sestavování dále uvedených schémat synt. analýzy, jde tu jen o vztahy těchto schémat).

¹⁵⁾ Srov. O. S. Kulagina, *O teoreticko-informacionnom podchode k izučeniju teoreticko-množestvennyh modelej jazyka*, MPPL 3, 1959, 95.

¹⁶⁾ Vedle grafické reprezentace větné stavby existuje ovšem mnoho jiných způsobů její reprezentace; tak s pomocí matic, různé způsoby lineárního zápisu s pomocí různých pomocných symbolů (závorek, svorek i běžných interpunkčních znamének apod.). Je důležité zjišťovat vztahy těchto různých reprezentací, protože mohou být nestejně vhodné pro jisté účely. — Z hlediska jisté algebraické reprezentace i reprezentace grafické vystupuje obecně vztah obou koncepcí Y. Lecker, *Une représentation algébrique de la structure des phrases dans diverses langues naturelles*, Comptes rendus heb. des séances de l'Acad. des sciences (France) 252, 1961, č. 2, 232—234. Výše uvedené pojetí vztahu obou koncepcí, k němuž jsme dospěli s B. Paltem při přípravě sdělení *Uplatnění teorie grafů v syntaxi přirozených jazyků* (pro seminář o teorii grafů pořádaný v Liblicích v květnu 1961 Matematickým ústavem ČSAV), je zhruba v souladu s Leckerovou prací. Srov. dále zejména S. Ja. Fitalov, op. cit. — K různým reprezentacím srov. W. Plath, *Automatic Sentence Diagramming*, Tedd., Paper 2.

¹⁷⁾ Odhlédneme tu tedy od otázek, nakolik mohou při synt. analýze pomoci právě lexikální a frazeologické údaje.

¹⁸⁾ K terminu S-struktura srov. A. Jaurisová — M. Jauris, *Užití teorie množin v jazykovědě*, SaS 21, 1960, 35.

¹⁹⁾ Synt. třídy mohou, ale nemusí fungovat jako synt. charakteristiky, viz 3.

²⁰⁾ Viz K. E. Harper — D. G. Hays, *The Use of Machines in the Construction of a Grammar and Computer Program for Structural Analysis*, Inf. Process., 191.

3.2. Mějme nyní seznam synt. tříd sl. tvarů pro jistý jazyk. Necht jsou X, Z (popř. s indexy) proměnné, oborem jejichž hodnot jsou symboly těchto tříd. Každý řádek v seznamu konfigurací může mít tento tvar

$$k \quad X_i X_j : Z \quad (4)$$

kde k — symbol (pořadové číslo apod.) konfigurace, Z — synt. třída, která má — z hlediska daného úkolu formální synt. analýzy — stejné synt. (kombinační) vlastnosti jako dvojice $X_i X_j$, a buď $Z = X_i$ (to znamená, že X_i je řídicím a X_j závislým členem konfigurace), nebo $Z = X_j$ (to znamená situaci opačnou).²¹⁾ Symbolům před dvojtečkou budeme říkat levá strana řádky, symbolům za dvojtečkou pravá strana. Tento způsob zápisu konfigurací se může zdát neusporný, avšak má tu výhodu, že je patrna souvislost s tvarem (5), jehož je tvar (4) jen speciálním případem. Obecně totiž nemusí být $Z = X_i$ nebo $Z = X_j$; k tomu, aby se zaznamenalo, že je řídicím členem X_i nebo X_j , stačí označit jeden z členů konfigurace (dále vždy člen řídicí) nějakým indexem, např. apostroфом:

$$k \quad X_i X'_j : Z \quad (5)$$

Přitom předpokládáme, že antepoňované nebo postpoňované postavení členů konfigurace je udáno pořadím symbolů v řádku na levé straně řádku tvaru (4), (5) nebo (6). Závažnost postavení začátkového, středového nebo koncového některého z členů konfigurace je někdy účelně vyznačit zvlášť (srov. pozn. 31). Pak ovšem dostaneme situaci poněkud složitější, než jaká je dále zachycena. Tvar (5) je nejobecnější způsob zápisu.²²⁾

Jak jsme již uvedli v 2.2, je dvojitelnost vztahu synt. závislosti podstatnou složkou koncepce A. Je to vidět i na tom, že z tohoto hlediska nezáleží na počtu členů konfigurace. Teoreticky pro jakékoli přirozené číslo n zápis tvaru

$$k \quad X_{i_1}, X_{i_2} \dots X_{i_n} : Z \quad (6)$$

kde je právě jen jedno X označeno apostroфом, znamená, že všechna ostatní X jsou na tomto X' jednotlivě závislá. Kvůli jednoduchosti však budeme dále většinou pracovat jen s dvojitelnými konfiguracemi. Pro další účely budeme skutečnost, že X je nebo není označeno apostroфом (tj. funkci členu X v konfiguraci), označovat symbolem f .

3.3 Máme tedy pět prvků (k, X_i, X_j, Z, f), z nichž můžeme vytvářet různé kombinace.²³⁾ Některé z těchto kombinací představují typ synt. charakteristiky, jaký lze v jistém schématu synt. analýzy přiřadit konkrétnímu sl. tvaru. Např. v jednom takovém schématu je synt. charakteristika typu kf , což znamená, že sl. tvar má v konfiguraci k jistou synt. funkci (že je např. řídicím členem). To je však jen jedna z dvou možných interpretací, a to interpretace „přímá“. Všech 23 kombinací můžeme však interpretovat také jinak, „nepřímou“, s pomocí druhé

²¹⁾ K symbolice srov. S. M. Lamb, *On the Mechanization of Syntactic Analysis*, Teadd., Paper 21, s. 5.

²²⁾ Jsou však možné i jiné způsoby, srov. např. symboliku T. N. Moloně *Algoritm perevoda s anglijského jazyka na ruskij*, PK 3, 209—272, zvl. 253n., kde je řídicím členem konfigurace ten, jehož tvar symbolu má jistý vztah k tvaru symbolu Z . Např. konfigurace $2+P_m:2$ ($2+$ — přechodné osobní sloveso, P_m — osobní zájmeno v předmětovém pádu, $2'$ — neprechodné osobní sloveso), kde řídicím členem je $2+$.

²³⁾ Celkem 31 kombinací 1. až 5. třídy. Z takto vytvořených kombinací můžeme pak pro naše účely ztotožnit ty, které se od sebe navzájem liší jen tím, že je v jedné X_i a v druhé X_j , tedy např. $kX_i = kX_j$. Jak se lze snadno přesvědčit, zbudě celkem 23 kombinací, totiž 4 kombinace 1. třídy, 7 kombinací 2. třídy, 7 kombinací 3. třídy, 4 kombinace 4. třídy a 1 kombinace 5. třídy.

ho členu konfigurace. V našem případě (kf) by tedy šlo o to, že se sl. tvar vyskytuje v kombinaci k , jejíž druhý člen je členem řídícím.

3.3.1. Některé z těchto 46 (tj. 23×2) kombinací se v úloze synt. charakteristik v literatuře již vyskytly, totiž $X; kf; X_i X_j$ (všechny tři v „přímé“ interpretaci), většina nikoli. Je nyní otázka, zdali bylo použito již všech možných kombinací. Ukážeme si dále na jednom případě (X/Z v nepřímé interpretaci), že nikoli.

Bylo by ovšem třeba nejprve prozkoumat všechny zbývající kombinace, dále pak zjistit pravidla vymezení kombinace možné proti kombinacím nemožným. Součástí těchto pravidel bude jistě zjištění, které kombinace a za jakých podmínek určují jistou synt. třídu a dále — vzhledem k tomu, že výběr kombinace má vliv na tvar řádku v seznamu konfigurací — i zjištění souvislosti s různými typy redukce tohoto seznamu. Nabízejí se ovšem i otázky další, např. zda je možné, že výběr některých kombinací a z toho vyplývající úpravy seznamu konfigurací spolu s jistým typem postupu při synt. analýze budou pro některé jazyky vhodnější než kombinace jiné nebo pro jiné jazyky. Seznamy konfigurací se od sebe mohou lišit vedle rozsahu (počtu řádků) a tvaru řádku i tím, že mohou být uspořádané nebo neuspořádané.

3.4 Klasifikace, jakou jsme vlastně navrhli v 3.3 pro synt. charakteristiky, je ovšem jen součástí budoucí celkové klasifikace schémat synt. analýzy.

O postupu synt. analýzy lze říci zatím tolik; je omezen dvěma momenty: jednak musí zabezpečit zjišťování synt. závislosti nejen mezi sousedními sl. tvary, jednak je jasné, že se tak nemůže dít prostě přezkoumáním všech kombinací 2. třídy ze symbolů vyskytujících se v dané S-struktuře, že se tedy musí počítat s jistými slovoslednými pravidly. Obvykle se začíná se zjišťováním syntaktických závislostí mezi sousedními sl. tvary. Uvádějí se různá pravidla omezující hledání závislosti mezi sl. tvary. Jedno z nich se dá formulovat takto:²⁴⁾ v zápisu vnitřní struktury s pomocí závorek — všechny sl. tvary závislé na jistém sl. tvaru jsou i se sl. tvary závislými na nich ve společných závorkách (včetně sl. tvaru řídicího), přičemž členy závislé mají po jedné závorkách navíc, např. ((Otec dočítel (dlouhý) dopis)) — nemůže dojít k případu:

$$(k (l))k \quad (7)$$

kde k a l označují závorky, které „k sobě patří“.

Jak uvidíme, je možný postup „od textu k systému“, a to buď „zleva doprava“ (viz 3.512), nebo „zprava doleva“ (viz 3.521), i postup „od systému k textu“ (viz 3.511).^{24a)} Často se uplatňuje postupná redukce S-struktury (viz 3.511, 3.521, 3.53), vedle ní však i postupy jiné (viz 3.512, 3.522).

Dále bude třeba zkoumat další otázky, především vztah možných synt. charakteristik, možných úprav seznamu konfigurací a možných typů postupu. Např. postup od „systému k textu“ není možný u synt. charakteristiky tvaru X/YZ a $X_i f X_j$.

3.5. Provedeme nyní ukázkový rozbor jednoduché anglické věty *Dunn chooses electrical analogs*. Větu anglickou jsme zvolili s ohledem na to, že většina schémat byla předvedena právě na anglickém materiálu.

3.5.1. *Syntaktická charakteristika tvaru X*. Jako synt. charakteristika funguje přímo symbol synt. třídy. Mohli bychom tu mluvit o symbolice „značkové“ (v Mathesiově smyslu). Nejde o to,

²⁴⁾ K. E. Harper — D. G. Hays, l. c.; S. Ja. Fitalov, op. cit., 5; srov. i jiné pravidlo u G. S. Cejtina a L. N. Zasořinové *O vjdelenii konfiguracij v ruskom predložení*, Dokl. 61, vyp. 2, s. 9. — Viz dále W. Plath, *Automatic Sentence Diagramming*, Teadd., Paper 2, s. 7n.

^{24a)} I tu jsou možné různé modifikace.

že by symboly pro jednotlivé synt. třídy nemohly být jistým způsobem popisné — a bývá to i dokonce výhodné (srov. pozn. 22) — nýbrž o to, že je tu z obecného hlediska jakákoli popisnost symbolů nepodstatná. Tato symbolika předpokládá úplný soupis konfigurací. Pro analýzu S-struktury je možno postupovat dvojným způsobem: buď vycházet při analýze z uspořádaného souboru konfigurací (ze „systému“), nebo z „textu“ (pak je množina konfigurací neuspořádaná).

3.511. „Systém-text“. Stroj přiřadí jednotlivým sl. tvarům analyzovanou větu symboly synt. tříd.²⁵ V našem příkladě (vpravo dole je u každého symbolu pořadové číslo sl. tvaru).

Dunn	chooses	electrical	analog	(8)
1 ₁	2+2	3 ₃	1 ₄	(9)

Ve stroji je uložen v jistém pořadí seznam konfigurací.²⁷ V seznamu jsou i tyto řádky (tečky označují jistý počet řádek, na nichž nám právě nezáleží):

3 1	→	1
2+1	→	2

Operace na S-struktuře se provádějí takto: Zjistí se, vyskytuje-li se v ní posloupnost totožná s posloupností symbolů levé strany v 1. řádce seznamu konfigurací. V kladném případě se do S-struktury dosadí za všechny případy takové posloupnosti symboly z pravé strany 1. řádky. V záporném případě nebo po vykonání popsané operace se přistoují k obdobné konfrontaci 2. řádky s posloupností symbolů již upravenou atd. V našem případě bude první úprava mít tvar 1₁2₂+1₄, čímž se zjistí, že sl. tvary *electrical* a *analog*s tvoří konfiguraci s *analog*s jako řídicím členem (podle zápisu, kterého budeme dále užívat, *electrical* ← *analog*s — šipka směřuje k závislému členu) a druhá (a konečná) úprava bude mít tvar 1₁ 2₂, čímž se zjistí, že jednak *chooses* → *analog*s, jednak *Dunn* ← *chooses* („predikační“ konfigurace).

3.512. „Text-systém“ (viz K. E. Harper — D. G. Hays, op. cit., 191). V Harperově systému jsou symbolika a soupis konfigurací podobné jako u Mološné, s tím rozdílem, jak již bylo poznamenáno, že na pořadí řádků v seznamu konfigurací nezáleží. Postup je tento. Mezi symboly synt. tříd se vsune symbol *p*, v našem případě tedy 1₁ *p* 2₂ *p* 3₃ 1₄. Postupuje se dále zleva doprava. První levá dvojice 1₁ *p* 2₂ se srovná se seznamem konfigurací a podle výsledku hledání se na zvláštní místo zapíše buď 1₁ *d* 2₂+2 (závislí-li 1₁ na 2₂), nebo 2₂+*d* 1₁ (závislí-li 2₂ na 1₁),²⁸ v našem případě tedy 1₁ *d* 2₂+2. Další postup se řídí ve shodě s těmito zása-

²⁵ Budeme postupovat podle systému T. N. Mološné *Nekotorije voprosy sintaksisa v svjazi s masinym perevodom s anglijskogo jazyka na russkij*, Vjaz 1957, č. 4, 92—97, *Algoritm perevoda s anglijskogo jazyka na russkij*, PK 3, 1960, 209—272. — K tomuto typu v zásadě patří velký počet publikovaných přístupů: tak I. K. Bel'skaja, *Machine Translation Methods and Their Application to an Anglo-Russian Scheme*, Inf. Process., 199—217; G. H. Matthews — S. Rogovin, *German Sentence Recognition*, Mechanical Translation 5, 1958, č. 3, 114—120; N. D. Andrejev, op. cit., kap. 5; T. M. Nikolaeva, *Struktura algoritma grammatičeskogo analiza (pri MP s russkogo jazyka)*, MPPL 5, 1961, 27—44. V některých se začíná od členů závislých, v jiných od členů řídicích.

²⁶ Zvolili jsme příklad tak jednoduchý, že je možno se synt. homonymií, tj. v podstatě s přirazením jednomu sl. tvaru více než jedné synt. charakteristiky, a s jejím odstraněním, tj. se zjišťováním, která charakteristika je jediné vhodná pro danou konkrétní větu, téměř nepočítat. — V 3.521 jsme toho dosáhli tím, že udáváme jen konfigurace č. 7 a 12, avšak v S-struktuře (20) jsme museli pro jednoduchost např. sl. tvaru *electrical* přiřadit arbitrárně charakteristiku (*n/n**), ačkoli bychom mu měli přiřadit i charakteristiku (*n/n*) a ještě některé další.

²⁷ Srov. pozn. 22. Apostrův řídicího členu je tu suplován příbuzností tvaru symbolu řídicího členu k symbolu z pravé strany řádky.

²⁸ Harper-Hays, op. cit., 191: „... jsou-li X, Y symboly ve větě, XpY znamená, že X předchází Y, a XdY znamená, že X závisí na Y“. A dále: „Jestliže WdXd... dYdZ, říkáme, že je W „odvozeno z“, X...Y a Z; základní symbol X nezavisí na sobě samém, ale každé konkrétní X je ze sebe odvozeno.“ (Překlad není zcela věrný, v originálu jde při X, Y apod. o konkrétní sl. tvary (occurrences), zde již o symboly synt. tříd.)

dami: a) XpY se zavádí, je-li pořadové číslo symbolu X v S-struktuře menší než pořadové číslo symbolu Y, jestliže každý symbol, jehož pořadové číslo je větší než X a menší než Y, je odvozen buď z X, nebo z Y, jestliže ani X, ani Y nejsou od sebe vzájemně odvozeny a jestliže X, Y nebo obě jsou nezávislé; b) XdY se zavádí, jestliže X a Y jsou uvedeny v seznamu konfigurací, jestliže XpY nebo YpX a jestliže nebyla předtím zjištěna závislost X na jiném symbolu.

Další dvojice, tj. 2+2 *p* 3₃ se v seznamu konfigurací nenajde, avšak dvojice 3₃ *p* 1₄ ano, proto 3₃ *d* 1₄. Zjistili jsme tedy zatím, že *Dunn* ← *chooses* a *electrical* ← *analog*s. Podle zásady a) je však třeba ještě zavést 2+2 *p* 1₄, načež hledání v seznamu konfigurací dá 1₁ *d* 2₂+2 (tj. *chooses* → *analog*s). Tím je synt. analýza věty u konce.²⁹

3.52. Při dalších tvarech synt. charakteristiky můžeme mluvit o symbolice „popisné“. 3.521. *Synt. charakteristika* tvaru kf. V systému cambridžské skupiny³⁰ se každému sl. tvaru přiřazuje symbol konfigurace (symboly konfigurací), v níž (v nichž) se může vyskytnout, jakož i symbol možné synt. funkce (člen řídicí, závislý, obojí). Do paměti stroje již proto není nutno ukládat seznam konfigurací celý, nýbrž jen seznam symbolů konfigurací, jimž jsou opět přiřazeny symboly téhož druhu jako sl. tvarům [viz zde (11)].

Pro náš případ:³¹

Dunn	(7)	(12)	(7)	(12)
chooses	2	1	12	0
electrical	0	2	7	3
analog	1	0	3	1
	2	1		

Postupuje se takto (opět zjednodušujeme): Zjistí se, zda mohou být poslední 2 nebo 3 sl. tvary členy 1. konfigurace, 2. konfigurace atd. V případě kladném se symboly přiřazené oběma (všem třem) sl. tvarům nahradí symboly přiřazenými příslušné konfiguraci a postup se opakuje. V případě záporném se obdobně vyšetřují dva (tři) sl. tvary od konce.

V našem případě se tak nejdříve zjistí, že *electrical* a *analog*s mohou být závislým, resp. řídicím členem 7. konfigurace. Jejich řádky se proto nahradí řádkem přiřazeným 7. konfiguraci [v (11)], tedy

(7)	(12)
2	1
0	2
3	1

a v dalším se zjistí, že *Dunn* ← *chooses* → *analog*s, čímž je synt. analýza u konce — všechny tři řádky nahrazeny řádkou přiřazenou 12. konfiguraci.

3.522. *Synt. charakteristika tvaru XifX*. Notací variantou tohoto tvaru je synt. charakteristika tvaru Y/Y', kde Y' je symbolem pasivní valentnosti (schopnosti dané synt. třídy být závislá na jisté jiné synt. třídě, třídách), Y symbolem aktivní valentnosti (schopnosti dané synt. třídy být

²⁹ Jiným typem postupu „text-systém“, a to „zleva doprava“, je patrně tzv. prediktivní synt. analýza, viz např. M. E. Sherry, *The Identification of Nested Structures in Predictive Syntactic Analysis*, Tedd., Paper 18.

³⁰ Viz A. F. Parker-Rhodes, R. Mc Kinnon Wood, M. Kay, P. Bratley, *The Cambridge Language Research Unit Computer Program for Syntactic Analysis*, ML 136, 1—86.

³¹ Symbol 0 v řádce u jistého sl. tvaru a v sloupci *n* znamená, že sl. tvar se v konfiguraci *n* vyskytnout nemůže, 1 — že se v ní může vyskytnout jen jako člen závislý, 2 — jako člen řídicí, 3 — jako obojí. Pro nás jednoduchý případ stačí údaje o 7. (nominální) a 12. („free clauses“) konfiguraci. Jinak silně zjednodušujeme. Systém cambridžské skupiny má ve skutečnosti 12 konfigurací, k synt. charakteristice sl. tvaru patří dále symbol možné jeho pozice v dané konfiguraci a údaje o shodě.

NEKOTORE PROBLEMY SINTAKSICKESKOGO ANALIZA S TOČKI ZRENIA MAŠINNOGO PEREVODA

Syntaktický analýza v celách mašinného prevodu javlja sa jednou z zadač analýtického charakteru. Ot takých (i iných) zadač, stavajučich prevodnaja neodnakoju strósti, a inogda i neodnakoju celenapravlenosti k našemu poznanju jazyka, sledujuč otličat s odnoju storony samo opisanje, kotore doľno zadovolvortja etim prevodnjam, s drugoju storony opredelenne obščie predstavlenja o strókture jazyka (ili jazyka kak celogo, ili že kakoj — nbnudj ego časti), nachoďne svoje primenenie pri opisanii i, nakonec, opredelenne jazykovoje ili vobščie simvoličeskoe obrazovanije, vypolnjujuče opredelenuju zadaču i opirajuče na adzkatnoe opisanie, v kotoremu provljaetja opredelennaja koncepcija. Iz syntaktickeskojo analiza v celach mašinnojo prevoda v dannoj statje rasmatrivajetja etich otnošenij v obščem osnovanicheskich otnošenij. Schemy algoritmov ustanovlenija etich otnošenij A — vključajuče na dnuh izvestnych koncepcijach, zdesj nazyvajemych prosto koncepcij A — vključajuče v osnovno obščie čerty tradicijonnoj koncepcii o členač predloženia i koncepcii o slovosocetačnjach — i koncepcij B — sootvetstvujučej, v osnovno, t. naz. analizu po neposredstvenno sostavljajučim.

V statje rassmatrivajetja boľe podrobno koncepcija A. Iščoda iz nee nužno prič ustanovlenii syntaktickeskich otnošenij ustanovitja dlia každoj zavisimoj slovoformy sootvetstvujučuju upravljajučuju slovoformu. Dlia togo čtoby sozdatj algoritm, vypolnjujuč etju zadaču, neobčodimo a) obladatj perečnem opredelennych syntaktickeskich klassov slovoform, b) perečnem elementarnych syntaktickeskich konstrucij. t. naz. konfiguracij (v terminologii syntaktickeskich klassov), v) izvestnymi poznanijami o porjadke slov.

Syntaktickesкий анализ проводится не непосредственно на предложении, а на последовательности индексов, представляющих собой определенные синтаксические данные об отдельных словоформах. В статье ставится вопрос, чем могут быть представлены синтаксические данные о словоформах, если исходить из концепции А. Строчки в (б) могут иметь форму (5), где k — символ конфигурации, X, Z — символы синтаксических классов, — обозначение управляющего члена конфигурации; (5) следовательно обозначает: в конфигурации $k - X_i$ является зависимым, а X_j управляющим членом, вся конфигурация относится к синтаксическому классу Z . Функцию члена X , обозначенную присутствием/отсутствием зн. ' будем обозначать символом f . Следовательно у нас имеется 5 элементов (k, X_i, X_j, Z, f), из которых мы можем составить всего 31 комбинацию от 1 до 5 классов. Некоторые из этих комбинаций представляют собой форму синтаксических данных, которые уже использованы, некоторые нет, а некоторые мам. Некоторые комбинации были уже использованы, некоторые нет, а некоторые невозможны в качестве синтаксических данных. Часть 3. 1—3. 4 является, в сущности, проектом частичной классификации схем алгоритмов синтаксического анализа для целей машинного перевода (классификация будет полной, когда и остальные части синтаксического анализа будут подобным образом рассмотрены).

Подобная проблематика возникает и для концепции Б и вообще для любой другой синтаксической концепции: необходимо установить все возможные схемы приемов, выполняющих определенную задачу и исходящих из определенной концепции.

4.2. Ke klasifikaci obdobné té, kterou jsme uvedli v 3.3 a problematice s ní spojené, přistupuje tu klasifikace, jejímž dělícím principem je interpretace symbolu Z . Podle toho jsou možné dvě varianty koncepce B :

B_1 — v každém řádku tvaru (4) je Z symbolem syntaktické třídy, která se vyskytuje aspoň v jedné S-struktuře; $B_2 - Z$ takovým symbolem být nemusí. Příklady na jednotlivé varianty: B_1 — množinová koncepce O. S. Kulaševiča³⁵, B_2 — větně strukturální gramatika (phrase structure grammar) N. Chomského³⁶ a soustava $Y. Yngveholma$ ³⁷.

4.3. Z nich ovšem jen poslední dvě jsou zaměřeny k synt. analýze pro účely SP (Bar-Hillel má však náročnější cíl, srov. 3.53), kdežto koncepce Kulaginová má nealgoritmický charakter³⁸) a větně strukturální gramatika je v pojetí Chomského jedním z typů tzv. generativní gramatiky, jejíž úkol je syntetickým prostředkem úkolu (II) z I.2, totiž produkce jen gramaticky správně utvořených vět daného jazyka (a žádných jiných posloupností sl. tvarů).³⁹

Matematicky byla již dokázána ekvivalenčnost kategoriální a větně strukturální gramatiky (Bar-Hillel, op. cit., 8n.) a přesvědčivě zjištěna i ekvivalenčnost kategoriální gramatiky a množinového modelu Kulaginové (I. I. Revzin, op. cit. v pozn. 33).

5. Závěr. V 3.31 jsme vlastně naznačili problém, jak zjistit možné postupy (algoritmy) založené na koncepci A. Tým problém je ovšem i pro koncepci B a bylo by užitečné stavět jej pro jakoukoli takovou koncepci. Dále by bylo třeba obdobně probrat odstraňování homonymie a členění textu (srov. I.1) a jejich vztahy.⁴⁰

Pro podobná zkoumání je ovšem třeba — jak jsme se zde o to pokusili — nalézt hledisko dostatečně obecné, z něhož by bylo možno jednotně přehlednout různé práce, jejichž skutečné vztahy — názorové shody a rozdíly — jsou zatím terminologickou a pojmovou nejednotností.

³⁵) Ob odnom sposobe opredelenija grammatičeskich ponjatij na baze teorii množestv. PK 1, 1958, 203—214, u nás dobře známá z výkladů A. a M. Jaurisových, op. cit. v pozn. 18. Viz i poněkud odlišné výklady I. I. Revzina O nekotorych ponjatijach tak nazyvajučemoj teoretičko-množestvennoj koncepcii jazyka. Vjaz 9, 1960, č. 6, 88—94 a další rozvíjí I. I. Revzin, op. cit. v pozn. 11.

³⁶) Např. Syntactic Structures, kap. 4; op. cit. v pozn. 34; Y. Bar-Hillel, op. cit. v pozn. 32. S. Scheinberg, Note on the Boolean Properties of Context Free Languages, Information and Control 3, 1960, 372—5.

³⁷) Op. cit. v pozn. 4, s. 297—304.

³⁸) Srov. A. A. Zinov'jev - I. I. Revzin, Logičeskaja model' kak sredstvo naučnogo isledovanija, Vopr. filosofii 14, 1960, č. 1, 87.

³⁹) Zjednodušenou variantou synt. analýzy podle koncepce B je i systém A. D. Stathakopoulose A Possible Application of Electronic Computers to the Block Analysis of Greek Sentences, Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English, Part II, Seattle 1960, 449—474.

⁴⁰) Přestože byly o použitelnosti obou koncepcí pro účely SP vysloveny značně skeptické názory, viz zejména Y. Bar-Hillel, Some Linguistic Obstacles to Machine Translation, 2^e congrès intern. de cybern. Namure, 3—10 Sept. 1958, Namure 1960, M. Corbe - R. Tabory, Introduction to an Automatic English Syntax (by Fragmentation), Tedd., Paper 35, je založena většina známých algoritmů právě na nich. — Problematiku tzv. transformační analýzy probíráme při jiné příležitosti.