ČASOPIS PRO MODERNÍ FILOLOGII 97, 2015, Č. 1, S. 9–20

Si představ, by mě zajímalo, ti to ukážu: Jednoslabičné začátky syntaktických segmentů v mluvené češtině[1]

Jana Hoffmannová (Praha) – Olga Richterová (Praha)

ONE-SYLLABLE WORDS AT THE BEGINNINGS OF SYNTACTIC SEGMENTS IN SPOKEN CZECH
The paper deals with a phenomenon frequently encountered in the syntax of spoken Czech, namely one-syllable words, mostly of pronominal or verbal nature (*se, si, sem, ste, sme, mě, mi, mu, tě, ti, bych, bys, by…*) at the beginning of syntactic segments. At this stage, the analysis focuses on three forms: *by, si, ti*. The authors address the issue of the difficult identification of segment boundaries, including the influence of turn-taking in dialogue. The data was taken from the ORAL2013 corpus; the paper further looks into the usefulness of this corpus for the investigation of dialogue syntax, its query options and the possible interpretation of the presented evidence. The results have shown so far that the one-syllable beginnings in question are based on the elision of certain, mostly pronominal, expressions, or less frequently on word-order inversion. Furthermore, to a certain extent, they correlate with selected non-verbal discourse phenomena (longer pauses, silence, laughter), with syntactic phenomena (repetitions, corrections, parentheses, aposiopesis, etc.) and also with speaker turn-taking and topic change.

KEYWORDS
spoken Czech, spoken corpora, corpus annotation, speech segmentation, one-syllable words at the beginning of utterances / turns in dialogue

KLÍČOVÁ SLOVA
mluvená čeština, mluvené korpusy, anotace, syntaktická segmentace, jednoslabičné začátky výpovědí / dialogických replik

1. ÚVOD A TEORETICKÁ VÝCHODISKA

Hlavním tématem příspěvku je zamyšlení nad jednoslabičnými začátky syntaktických segmentů v mluvené češtině, založené zatím na analýze tří výrazů typických pro tuto pozici (***si*** *nemysli,* ***ti*** *řeknu,* ***by*** *mě zajímalo*). Motivací pro tento výzkum nám byla především vysoká frekvence jevu v mluvených projevech i v projevech psaných, výrazně ovlivněných mluveností (soukromé dopisy, maily, chaty, beletristická stylizace).

Hodnocení tohoto typu počátkových výrazů v odborné literatuře je značně rozrůzněné — zejména v otázce, zda tyto výrazy považovat za klitika. Např. J. Toman (2001, s. 78) formuluje pravidlo: „V současné spisovné češtině nemůže stát klitika

1 Tato studie vznikla v rámci Programu rozvoje vědních oblastí na Univerzitě Karlově č. P11 *Český národní korpus*, podprogram *Český národní korpus* a v rámci projektu GA ČR č. 406/12/1829 (*Stylistika mluvené a psané češtiny*).

 © FILOZOFICKÁ FAKULTA, UNIVERZITA KARLOVA V PRAZE

v absolutním větném začátku". Data, z nichž vycházíme, však většinou nejsou založena na spisovné češtině; a navíc je pro sledovaný jev centrálním problémem vymezení „absolutního větného začátku". Další autoři, mj. Sgall a Hronek (1992, s. 64–65) nebo M. Nekula (Karlík — Nekula — Rusínová, 1995, s. 648) jednoslabičné začátky chápou jako příklonky k výrazům, které by se na počátku výpovědi objevily v psaném spisovném projevu a v projevu mluveném bývají leckdy elidovány; vypuštěno může být podmětové zájmeno jako *já, to* (*sem tam nebyl; si něco jedla?*) nebo jiný typ výrazu, např. částice (*vždyť, tak* aj.: *Se mně zdá, že…* = ***Tak*** *se mně zdá, že…*). Jindy se jako motivace nepřízvučných jednoslabičných začátků uvádí hlavně slovosledná inverze, slovosledná varianta bezpříznakového postavení příklonek; podle A. Svobody (2000, s. 158) v případech jako *Se mu to nepovedlo* nejde o předklonky, ale o „inverzi iniciální a postiniciální slovosledné přihrádky", o příznakový, subjektivní, emotivní slovosled.

M. Nekula (Karlík — Nekula — Rusínová, 1995) uvádí, že některé z uvedených struktur lze interpretovat jako „polknutí předchozího výrazu" — tematizuje tedy v této souvislosti i zvuková specifika mluveného projevu. Toman (2001) vyzdvihuje u klitik těsný vztah mezi syntaxí a prozodií. Směřujeme-li při výkladu k příklonkám či předklonkám, nelze tu samozřejmě opomenout faktor rytmický; dalo by se uvažovat o „jambických" začátcích. To je ovšem možné jen u případů typu *ti řeknu, si myslim*; nikoli tam, kde před víceslabičným výrazem stojí dva výrazy jednoslabičné, což už rytmus podstatně mění: *ti to nešlo, si to užij…* Foneticky orientovaný výzkum by se tu měl zaměřit i na další zvukové aspekty: redukovaným začátkům velmi často předchází pauza (i delší), ale bylo by vhodné sledovat také roli počátečního rázu.

Analyzovaný problém je těsně spjat s nesnadnou identifikací začátků (a konců) syntaktických segmentů v mluvené řeči. Jde o obtíž poměrně bohatě zachycenou a reflektovanou v zahraniční i domácí literatuře (srov. např. Kvale, 1993; Brazil, 1995; Passonneau — Litman, 1996; Christiansen — Allen, 1997; Edlund et al., 2005; Rao et al., 2007; Müllerová, 1994 aj.; Čermák, 1996), ovšem je třeba potýkat se s ní stále znovu, především v důsledku rozdílů ve výstavbě a anotaci různých korpusů. Při výzkumu mluvené syntaxe neustále narážíme na problém stanovení jednotky mluveného projevu, neboť ostře ohraničené jednotky typu vět v psaném projevu zde nepřicházejí v úvahu. Neurčité, rozplývavé hranice jednotlivých úseků, útržkovitost mluvené syntaxe, časté dodávání a nastavování, juxtapoziční přiřazování, množství oprav a „falešných startů", ztrácení větné perspektivy — to vše segmentaci projevu komplikuje.[2] Důležitým signálem začátku segmentu je střídání mluvčích v dialogu, v korpusových datech ho lze ovšem využívat jen obezřetně; na hranicích replik ve spontánním živém rozhovoru dochází nezřídka k překrývání promluv, jeden mluvčí skáče druhému do řeči, „přebírá" od něj nedokončenou konstrukci. V tomto poněkud chaoticky strukturovaném prostoru vznikají slovosledné nepravidelnosti, eliptická vyjádření apod. Je proto na místě přistupovat k analýze tohoto a dalších podobných jevů z hlediska „dialogické syntaxe", z pohledu interaktivního, brát v úvahu i zvuková specifika a speciální techniky mluveného projevu — což právě současné korpusy, propojující transkripty s nahrávkami, ve vysoké míře umožňují.

Při manuálním segmentování korpusových dat lze tedy využívat tyto ukazatele:

2 Někteří autoři, např. Passonneau — Litman (1996), přímo doporučují pracovat s pojmem neostrých hranic — *fuzzy boundaries*.

a) obsahově-tematické uspořádání textu;
b) jeho organizaci z hlediska syntaktických konstrukcí;
c) jeho zvukovou stránku, tj. především pauzy, intonaci, ale i další jevy z oblasti prozodie, rovněž hezitační prvky;
d) prvky lexikální: Schwitalla (2012, s. 85 n.) uvádí mezi indikátory hranic v mluveném projevu např. částice plnící funkci strukturování a organizování hovoru, nebo krátké „počátkové" formule typu *ich meine, ich finde, ich glaube* (v češtině *si myslim* apod.);
e) výstavbu dialogu se zaměřením na „turn-taking" (střídání mluvčích).

Evidentně však neexistuje žádná „jediná správná" strategie segmentace mluvených projevů (srov. Kvale, 1993); hranice segmentů jako koherentních obsahově pragmatických jednotek se obtížně určují lidem i počítačům.

2. PRÁCE S KORPUSOVÝMI DATY

Tento příspěvek bezprostředně navazuje na předchozí studii J. Hoffmannové a I. Kolářové (2013), pro kterou byla využita hlavně data z korpusů SYN2000, SYN2005, SYN2010, ORAL2006 a ORAL2008. Nyní se nabízelo soubor analyzovaných dat rozšířit (kvantitativně i typově) a jako zdroj využít především velký korpus mluvené spontánní češtiny ORAL2013 zveřejněný na konci roku 2013. Obsahuje téměř tři miliony slov a nahrávky, na nichž se zakládá, byly pořízeny v Čechách, na Moravě i ve Slezsku.

2.1 VLASTNOSTI KORPUSU ORAL 2013

Pro účely zde prezentovaného výzkumu je podstatná zejména inovace webového přístupového rozhraní, která transkripci rozhovorů zahrnutých do korpusu ORAL2013 propojuje se zvukovou stopou. Neméně důležitý je však i přístup k transkripci a anotaci dat: způsob zpracování rozhoduje o tom, jaké jevy je možno jakým způsobem (a zda vůbec) vyhledávat. Významný posun proto představuje proznačování překryvů replik dvou mluvčích a vůbec možnost zobrazit střídání mluvčích v textu. Oproti předchozím korpusům řady ORAL se rozšířila též pauzová interpunkce rozlišující krátkou pauzu, delší pauzu a odmlčení. Jako nejdůležitější se nicméně ukázala jedna externě stanovená podmínka korpusu ORAL2013: Pravidla pro přepis stanoví, že žádný segment nesmí obsahovat víc než 15 slov. Ze situace, kdy značka <seg> spolehlivě indikuje začátek ucelené sekvence a </seg> její konec, se tak dostáváme do situace, kdy se <seg> vyskytuje jak u skutečných začátků syntaktických segmentů, tak v jejich středu, pokud byly prostě jen příliš dlouhé. Navzdory přítomnosti této úrovně v anotaci ji tak nebylo možno využít pro vyhledávání syntaktických začátků. Kromě toho zvukové stopy přiřazené k segmentům vykazují na počátku nepatrné zkrácení, které lze pravděpodobně pokládat za technický důsledek rozsegmentování; také tyto „překotné nástupy" ztěžují výzkum promluvových počátků.

Povaha dat zařazených do korpusu ORAL2013 je dialogická, promluv se účastní minimálně dva, často ale i více mluvčích. Vzhledem k neformálnosti a spontánnosti

zaznamenaných rozhovorů je časté vzájemné přerušování mluvčích, což dále ztěžuje identifikaci počátků syntaktických jednotek.

2.2 IDENTIFIKACE VLASTNOSTÍ DEFINUJÍCÍCH POČÁTEK SEGMENTU

Jak už bylo výše zmíněno, značku <seg> samotnou pro identifikaci počátku segmentu aplikovat nelze, neboť je v důsledku arbitrárního omezení délky segmentu na patnáct slov velmi nespolehlivým vodítkem. Pro členění mluvených dat jsou nicméně zásadní zvukové charakteristiky založené především na pauzách a přízvucích. Ani pauzy (*silent pauses, non-text*) ale nejsou samospasitelné, Edlund et al. (2005) upozorňují, že nehledě na význam pauz pro signalizaci prozodických hranic, spontánní promluvy často obsahují pauzy i v rámci takových jednotek, které ze syntaktického hlediska působí soudržným dojmem a které vykazují sémantickou koherenci. Schwitalla (2012, s. 76) dále upřesňuje, že je nutné rozlišovat dva typy pauz — ty, které přerušují, resp. ukončují promluvový celek (*Haltepause*), a ty, které ho pouze člení (*Auslaufpause*); ty první jsou dle autora delší. Spěje-li tematický celek k závěru, přibývá dle Schwitally pauz a kombinují se s dalšími prozodickými charakteristikami (např. hlas se stává tišším a hlubším) v jakési předtuše toho, že téma je vyčerpané a lze otevřít nové.

Na základě literatury a vlastního úsudku jsme za účelem identifikace skutečných počátků segmentů přidaly ke kritériu různě dlouhé tiché pauzy ještě tyto doprovodné nebo podpůrné zvukové charakteristiky: odmlčení, odkašlání, hezitační zvuky (*hmm, mmm, eee*), smích a neukončené výpovědi. Původně uvažované lexikální signály povahy částicové (*jo, no*) i vycpávkové výrazy (*takže, jako, jakože*) nakonec byly po úvodních sondách vyřazeny jako málo průkazné.

2.3 ZPŮSOB VYHLEDÁVÁNÍ V KORPUSU ORAL2013

Nejprve jsme vybraly tři jednoslabičné výrazy, na nichž byla založena korpusová sonda. Bylo zvoleno *by* jakožto slovesný element, *ti* jakožto osobní zájmeno a *si*, tj. výraz, který je realizací 2. os. sg. od slovesa *být* i zvratného zájmena.

Poté byl položen CQL dotaz prostřednictvím rozhraní KonText, a to v tomto znění: [word="eee|hmm|mmm|\.*|\(.*\)|\.\.\.\.:"] <seg> [word="si"] (s obměnou pro další dva tvary [word="ti"] a analogicky [word="by"]). Pauzy, které indikuje různý počet teček, jakož i neukončené výpovědi (značené třemi tečkami a dvojtečkou) a hezitační zvuky alternují ještě s komentáři v kulatých závorkách (např.: *odmlčení*). Značce počátku segmentu <seg> musel předcházet některý z těchto signálů, abychom se dané konkordanci dále věnovaly.

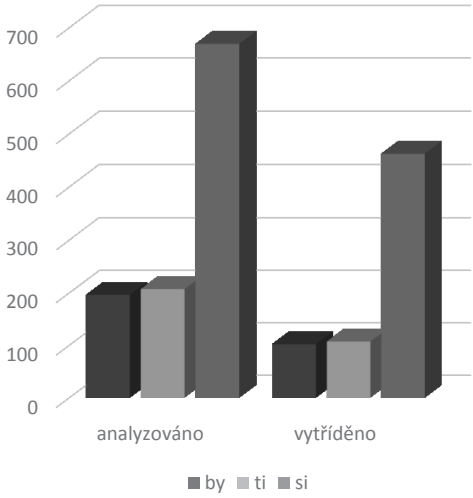
Příklad konkrétního postupu: dotaz vygeneroval 423 dokladů s *ti* v možné počáteční pozici. Obě autorky nezávisle prošly všechny konkordanční řádky a vyhodnocovaly, zda se jedná o zkoumaný jev (kvůli tvarové homonymii s plurálem demonstrativa, např. *ti chlapi*, se mezi konkordancemi nacházelo mnoho nežádoucích dokladů). Míra jejich vzájemné shody byla vysoká, ovšem nejméně u dvanácti konkordancí nebylo možno učinit rozhodnutí bez poslechu. Dále se ukázalo, že kratší pauza (značená jednou tečkou) není spolehlivým indikátorem, a přibližně třetina konkordancí byla patrně rozsegmentována arbitrárně, podle pravidla patnácti slov. Pauzy zřejmě ko-

relují s jednoslabičnými počátky jen do určité míry; v dalším výzkumu se asi vyplatí zohlednit zde rozdílnou délku pauz.

Z dalších vytipovaných zvukových signálů byly postupně zpochybněny i hezitační zvuky, které sice nesporně fungují jako obecný průvodní jev mluveného projevu, se zkoumanými počátky ale nekorelují (nízký souvýskyt je dán tím, že se tyto zvuky v našich datech nevyskytují tak často). Nezřídka plní čistě kontaktovou funkci: druhý mluvčí takovým zvukem promluvu vůbec nepřeruší, ale spíš povzbudí další řečovou produkci mluvčího dosavadního. Formu *hmm* tak lze s určitou nadsázkou označit za homonymní pro dvě rozdílné funkce: fatickou a výplňkovou (hezitační).

Naproti tomu jako klíčové se v průběhu analýz projevilo další kritérium: střídání mluvčích. To ale není možné vyhledávat automaticky — ač tuto informaci data implicitně obsahují[3] (u segmentů je uvedeno číslo hovořícího mluvčího), není v nich zanesena explicitně (jako samostatná značka): do dat získaných pomocí CQL dotazu musel být tento údaj doplněn manuálně. Využití tohoto kritéria je nezbytné, jinak by docházelo k chybnému vyhodnocování přepisů (*brou chuť </sp><sp>* ***děkuju****.. </sp><sp>* ***ti to je pohoda*** **—** bez indikovaného střídání pomocí značek <sp> by došlo k interpretaci celků jako „děkuju ti“ a „to je pohoda“)

2.4 VÝSLEDKY VYHLEDÁVÁNÍ



GRAF 1: Poměr analyzovaných a vytříděných konkordancí

Graf 1 shrnuje, kolik z analyzovaných konkordancí bylo identifikováno jako doklady námi zkoumaného jevu. Ukazuje, že až na výraz *si*, kde je poměr příznivější (69 % žádoucích konkordancí), výše uvedený CQL dotaz identifikoval zhruba polovinu do-

3 Nepomůže ani vyhledávání zdánlivého střídání mluvčích pomocí **</sp><sp**: končí-li segment, vyznačuje se takto, ale tentýž mluvčí může pokračovat: <sp num=02> <seg> vypadalo za dvacet let to čeк. </seg>]</sp><sp num=02> <seg> ti nikdy neví </seg> </sp>

kladů chybně (u *ti* i *by* správně 52 %). Určitý podíl na velkém počtu nežádoucích konkordancí má ovšem tvarová homonymie zvolených výrazů: u *ti* je pro sledované začátky relevantní osobní zájmeno ve 3. pádě sg. (*ti říkám*), nikoli demonstrativum v 1. pádě pl. (*ti chlapi*). Za jednoslabičné nelze pokládat ani začátky, kde počátkové *by* zastupuje nedbale artikulovaný výraz *aby*.

3. INTERPRETACE

Jako nejspolehlivější signály — a zároveň motivace, příčiny, byť i neuvědomělé — spojené s příznakovými jednoslabičnými začátky segmentů se osvědčily tyto tři: (1) střídání mluvčích, (2) elize určitého prvku pociťovaného jako obligatorní (nejčastěji zájmenného výrazu), případně (3) změna tématu.

(1) Vhodnou pozici pro příznakové jednoslabičné začátky v rozhovoru tedy vytvářejí především centra jeho interakční výstavby — momenty **střídání mluvčích**. Sledované jednoslabičné výrazy stojí často na počátku **repliky**, tam, kde se ujímá slova — nebo přímo přerušuje partnera, skáče mu do řeči — nový mluvčí. Ostatně např. M. Nekula (Karlík — Nekula — Rusínová, 1995, s. 647) zdůrazňuje, že v mluvených projevech je počátkové postavení nepřízvučných výrazů prostředkem navazování na předchozí repliku partnera v dialogu. Typické mezireplikové navazování je přitom navazování odpovědi na otázku, nezřídka právě prostřednictvím „jednoslabičné redukce“:

— *ten byl hodně dobrej? </sp><sp>* ***ti*** *ho někdy pučim. ten se ti bude fakt líbit*[4]
— *máte program do kolika. do kolikátého? (odmlčení) </sp><sp>* ***ti*** *nevim ani*

A kromě „našich“ příznakových počátkových výrazů *by, ti, si* se na počátcích, ev. i koncích replik poměrně hojně objevují také typické výrazy *jo, no, to* — které přes svou jednoslabičnost mají jinou povahu, ale plní také funkci navazovací, případně odkazovací (srov. o některých z nich Müllerová, 1996, 2007). Ukázkami chceme zároveň předvést, jak živé jsou někdy ve spontánním dialogu interakční výměny, jak rychle se mluvčí střídají (tomu zřejmě vyhovují redukované začátky):

— *máš hezký stíny. </sp><sp>* ***jo?*** *děkuju </sp><sp>* ***ti*** *ladí ke svetříku*
— *vaše jazykové znalosti sou na úrovni základní školy (odmlčení) </sp><sp>* ***ti*** *někdo říkal* ***jo?*** *</sp><sp>* ***jo to*** *mi řek von*
— *malej Peťa s Pavlem sou po mámě (odmlčení) </sp><sp>* ***si*** *to rozdělili* ***jo****? </sp><sp>* ***no****. malej Peťa* ***jo***

Na mimořádnou důležitost zvukové stránky při analýze mluveného projevu jsme zde už poukazovaly; v některých sporných případech nám také při identifikaci sledovaných začátků účinně napomohl **poslech** (umožnil přesné rozpoznání střídání mluvčích, překryvů jejich replik, povahy pauzy; přerušení pomohla identifikovat i intonace, apod.):

4 Tečka signalizuje krátkou pauzu, dvě tečky pauzu delší.

— *takže ti bylo kolik v roce třicet devět.. **ti** bylo. dvacet tři. třicet tři. ňákejch šestnáct* (poslech pomáhá identifikovat na základě intonace, že jde o začátek nové výpovědi)
— *že žádné materiály k tomu nejsou prý.. **ti** řeknou* (poslech identifikuje výrazné přerušení a nový začátek).

(2) Kromě předvedené motivace založené na replikové a sekvenční výstavbě dialogu hodnotíme příznakové jednoslabičné začátky syntaktických segmentů nejčastěji jako výsledek **eliptických** procesů, a to hlavně elize zájmen, případně některých výrazů částicové, příslovečné či spojkové povahy. V pozadí jednoslabičných začátků může být elize

a) zájmena ve funkci podmětu: *(to) **by** mě zajímalo, (já) **ti** dám vědět, (ty) **si** naštvaná, (voni) **si** to rozdělili*;[5] příznakové začátky spojené s elizí podmětu se mohou vyskytnout i víckrát za sebou, v celých sériích: *do toho ještě chlasta.. mmm (odmlčení) --- **se** úplně zabije.. **by** se měl trošku brzdit*
b) zájmena ve funkci předmětu: *(to) **si** představ, (to) **ti** řeknu*
c) částicového „to“: *(to) **by** byli blbý, (to) **si** děláš srandu*
d) dále též elize *tak (**si** zapamatujte), že (řek ti **si** dělník)* aj.

Výklad sledovaných začátků na základě eliptické redukce by nasvědčoval tomu, že zde jednoslabičné výrazy opravdu můžeme považovat za klitika, zbavená předchozího přízvučného slova. Elidované výrazy jsou zřejmě prostředkem kontextové zapojenosti, která funguje i po jejich vypuštění. Mnohdy ovšem nelze jednoznačně určit, jaký prvek byl elidován — na konci následujícího příkladu např. může být elidováno *tak* či *to*, případně i *vona*: *2: bych byl ještě svobodnej teďka — Y: (odmlčení) — 3: no — Y: (smích) — 3: to by bylo zajímavé — 1: **by** tě uhnala jiná*. Elize má někdy i větší rozsah, dochází i ke „spolknutí“ několika slov: *za to škubnul aby si. abys to vytočil. **ti** taky nepovede napoprvé že jo* (po pauze elize *vono se + to: vono se ti to taky nepovede napoprvé*).

V řadě případů je navíc nesnadné interpretační rozhodování mezi elizí a **slovosledn ou inverzí** (***si** to zapamatujte* = „tak si to zapamatujte“, nebo „zapamatujte si to“?). U ***ti** nevěřim* lze uvažovat buď o elizi „to“, nebo o slovosledné inverzi („nevěřim ti“); v mluveném projevu je ale „ti nevěřim“, resp. „to ti nevěřim“ přirozenější, tj. méně příznakové než „nevěřím ti“ (slovosled psaného projevu), proto preferujeme eliptickou interpretaci. Jednoslabičné začátky jako výsledek slovosledné inverze jsou ale jen jedním z projevů „neuspořádaného“ slovosledu mluvené řeči, který se nám ovšem tak jeví na pozadí organizace textů psaných. Z dat shromážděných při výzkumu začátků *by, ti, si* uvádíme aspoň jeden příklad markantního slovosledného rozlišení: ***si** dá siestu že?.. po obědě dycky dědeček* (proti „korektnímu“ slovosledu, jaký by bylo možno očekávat v psaném textu: *dědeček si dá po obědě vždycky siestu, že?*).

5 Nevyjádření podmětu je v češtině samozřejmě časté a gramaticky korektní; přesto elize podmětového zájmena v těchto případech působí příznakově, jako redukce, „spolknutí“ začátku.

(3) Jednoslabičné začátky se někdy pojí s náhlou **změnou tématu**, přechodem k tématu nebo subtématu jinému. Jasné případy tohoto typu jsou však spíše vzácné, např.: *na todle von byl zase rapl ten Kubrick.* ***si*** *pudu odskočit; tady je to (odmlčení)* ***si*** *taky skočím na záchod (odmlčení) </sp><sp> no tak si tam rozsviť.* Ve většině případů je obtížné rozlišit, kdy dochází k změně tématu a kdy jde jen o tematický posun či dílčí změnu v rámci společného hypertématu. Proto se u indikátorů (1) — (3) ještě neodvažujeme striktně kvantifikovat, ke kvantifikaci jejich souvýskytu s jednoslabičnými počátky ovšem v našem výzkumu směřujeme. Spolu s těmito třemi indikátory se pak jako pomocné indikátory počátku promluvových segmentů, podporující jednoslabičné začátky, uplatňují už zmíněné zvukové charakteristiky (různě dlouhé pauzy, neurčité hezitační zvuky, ale i „odmlčení" nebo „smích", zachycované v přepisech komentáři); jejich souvýskyt se sledovanými jednoslabičnými začátky není nijak výrazný — až na případy, kdy se vyskytuje pospolu několik indikátorů tohoto typu (např. pauza či odmlčení s hezitačním zvukem či s vycpávkou). Rovněž jednoslabičné výrazy se ale v počátkové pozici někdy kumulují — vznikají jejich „souběhy", z nich některé i relativně ustálené (***si mi to*** *řikala;* ***by ti to*** *neřek*).

V případě homonymního *si* je v počátkové pozici daleko frekventovanější reflexivní zájmeno (***si*** *dovedu představit*) než tvar 2. os. sg. slovesa *být* (***si*** *nemožná*). Někdy se sejdou i obě homonymní *si* bezprostředně za sebou: *hmm.* ***si*** *chytrý.* ***si*** *to peč sám.* K hledaným jednoslabičným začátkům počítáme v tomto případě oba homonymní tvary.

Kromě toho výzkum potvrdil, že jednoslabičné začátky souvisí s některými jevy povahy syntaktickostrukturní, typickými pro mluvené projevy, což je zejména

a) opakování: *já* ***dám*** *zítra vědět..* ***ti dám dám*** *vědět; voda* ***by*** *se udělala víš (odmlčení)* ***by*** *se vysekala jenom ta drážka na tu vodu*
b) opravy, upřesňování: *to pédéáčko* ***má. by mělo mít****. jedenáct ampérů aspoň; ňákou firmu která* ***má. ti to na dobírku pošle***
c) parenteze, u níž jednoslabičný začátek může následovat za vsuvkovým přerušením (*by ju někdo vyzvedl..* ***má tam spoustu přátel****.* ***si*** *dycky zařídí*), ale někdy může stát i na počátku samotné vsuvky (*já nevim jak se menuje.* ***si nezpomenu****. on je z Luk*)
d) ztráta výpovědní perspektivy, „falešný start": *tam já bych jako docela.* ***by*** *se mi to líbilo; boty který nevadí že si* ***ti*** *nevadí že si zničíš totálně.* Patří sem i případy, kdy mluvčí v průběhu lineárního přiřazování výrazů při postupné realizaci syntaktické konstrukce opustí strukturu, kterou začal realizovat, a přejde k realizaci struktury jiné přes určitý spojovací článek, ve kterém se dvě syntaktické konstrukce prostoupí, vzniká jejich překryv (srov. o tom Hoffmannová, 2001): *to by nemělo nechat žádný následky jako* ***bolesti zad nebo tak****.* ***by*** *tady z toho typu úrazů bejt vůbec nemělo*. Obě prostupující se struktury sdílejí úsek *bolesti zad nebo tak*, spojující článek neutralizuje kolizi 4. a 1. pádu, a po něm následuje konstrukce začínající *by*.
e) apoziopeze, parcelace aj.: *víš že jo to byl vole to byl* ***... ti*** *třeba dávaj různý. hamburgry. ee prostě k tomu;* ***si přišla ještě ... si ještě přišla*** *o výbornou fázi kdy Daniela vypadala že omdlí*

f) rozhraní mezi uvozovací větou a reprodukcí řeči: *von řikal **si** přiď **ti** pomůžu*; sem by bylo možno přiřadit i případy, kdy jednoslabičný výraz následuje po citaci partnerova vyjádření, např.: *vůbec nevim že sem zženštily ty bláho </sp><sp> („)zženštily(“) **ti** řekla?* „Začátkovost” jednoslabičného výrazu (*ti*) zde sice není nijak markantní, určitý syntaktický předěl po citátovém prvku tu však vzniká a na jeho pozadí je možno vidět elizi *že/to že* (*to že ti řekla?*). Citát z jiného dialogu, po němž následuje vystřídání mluvčích a výraz údivu nového mluvčího, může být i rozsáhlejší, jak dokazuje výměna replik uvedená již výše: *vaše jazykové znalosti jsou na úrovni základní školy (odmlčení) </sp><sp> **ti** někdo řikal jo? </sp><sp> jo to mi řek von*.

Celá řada syntaktických struktur tedy s jednoslabičnými začátky přirozeně koexistuje a tyto jevy příznačné pro syntax mluvené češtiny mohou být považovány za pomocné indikátory sledovaných začátků. Předěl mezi typy konstrukcí, které takto fungují, a mezi případy, kde už nemůžeme o sledovaných začátcích uvažovat, je samozřejmě velmi neostrý, což dokumentujeme aspoň frekventovanými infinitivními konstrukcemi: *mně to je tak blbý **ti dávat** ty peníze; akorát mi chybí třeba jako tak. **si jít** zatančit; nevim jestli bych měl čas. **si** fakt k tomu dycky večer **sednout**…* Před *ti, si* + infinitivem je určitý syntaktický předěl, někdy i pauza; předpoklad, že i zde tyto výrazy představují počátky syntaktických segmentů, podporují také případy, kdy má infinitivní konstrukce povahu volně připojeného dodatku: *dokud sem. eee opravdu neměla na ten dům čas. **si** ho užít*. Nicméně tu asi nejde o hledaný typ; nepřízvučné jednoslabičné výrazy zřejmě mají povahu klitik, nenastává tu elize, nedochází ke změně mluvčího ani změně tématu, nejistá je i slovosledná inverze… Infinitivní kondenzáty typu *Poslanci se rozhodli **si** zvýšit plat*; *nebude problém **mu** udělit vánoční prémii* vzbudily rovněž pozornost J. Tomana (2000, s. 163–165) při jeho hledání klitik v nekanonických pozicích (k nimž patří právě pozice iniciální). Podle něho u infinitivních konstrukcí jde o kanonickou pozici klitik, prozodicky licencovanou (přítomnost prozodicky nenulového materiálu před klitikem umožňuje proklitické frázování). Toman srovnává infinitivní struktury právě s naším typem nepřízvučných začátků (*se mi nezdá, bych si taky myslel*); u těch se ovšem rozhoduje pro výklad eliptický, totiž pro „výjimečnou možnost v jistých kontextech foneticky nerealizovat takřka prázdnou matrici syntakticko-sémantických příznaků, která je jinak foneticky realizována jako *to* (***to** se mi nezdá*; ***to** bych si taky myslel*)“ (Toman, 2000).

4. SHRNUTÍ

Jednoslabičné začátky lze tedy chápat jako výrazný rys organizace mluveného (zvláště dialogického) textu a jako jeden z prostředků charakteristické tendence mluvčích ke zkracování, vyjadřování úspornému a redukovanému — tendence, která se v tomto případě projevuje zejména na exponovaných místech výstavby dialogu, jako je střídání mluvčích či změna tématu. Potvrdilo se, že výzkum mluvených projevů, jejich syntaxe a textové výstavby musí zohledňovat dialogičnost a brát v úvahu interaktivní aspekty. Z toho je třeba vyvozovat závažné důsledky pro anotaci a přípravu korpusů

mluvené češtiny: je třeba výrazně odlišovat promluvy jednotlivých účastníků dialogu, značení mluvčích (<sp>) musí jasně ukazovat, zda se mluvčí změnil, nebo ne; a i když si technická omezení budou žádat členění na menší celky, značka pro „arbitrární" a „funkční" segment by se mohla lišit, a umožnit tak výzkum skutečně přítomných promluvových předělů.

5. VÝHLED

Lepší a systematičtější poznání syntaxe mluvené češtiny je bezpochyby naléhavým úkolem české lingvistiky; domníváme se, že ke splacení tohoto bohemistického dluhu může přispět i analýza symptomatického jevu, kterým jsme se zabývaly v tomto článku. Ve výzkumu je možno pokračovat mnoha směry:

a) Po uvedených třech výrazech (*by, si, ti*) je třeba prostřednictvím korpusu ORAL2013 prozkoumat i fungování dalších výrazů, které obdobně obsazují počátkovou pozici, tj. *se, sem, sme, ste, mě, mi, mu, tě, bych, bys…*
b) Pokud jde o funkci jednoslabičných začátků, zatím jsme nevěnovaly větší pozornost jejich polyfunkčnosti, tj. relevanci pro zavádění nové informace (nového tématu, subtématu) a zároveň pro navazování (kohezi vnitroreplikovou i mezireplikovou; týká se to např. výskytu, resp. elize odkazovacích výrazů *to, tak* na počátku segmentu).
c) Je třeba prohloubit postupně získávané poznatky o korelaci jednoslabičných začátků s jevy neverbální povahy, které se podílejí na segmentaci mluveného projevu: je např. „delší pauza" nebo „odmlčení" spolehlivějším indikátorem tohoto typu začátku než pouhá „pauza"? Nakolik se jako segmentační indikátor v této korelaci uplatňuje třeba „smích"?
d) Stálo by za to zaměřit se i na některá ustálená spojení (někdy frazémy), do nichž se počátkové výrazy zapojují a jež mohou rovněž sloužit jako pomocné indikátory. V případě *by* je to např. spojení *by mě zajímalo*; v případě *ti* spojení jako *ti řeknu/řikám, ti můžu říct, ti povim, ti nevim, ti (to) ukážu* aj. V případě *si* např. *si řikám, si pamatuju, si myslim, si připadám jako, si dám, si žije, si děláš srandu/legraci/prdel* a spojení imperativní *si dej, si představ, si vem, si dej bacha, si to zkus…* Uvádíme aspoň několik příkladů, na nichž je možno opět sledovat souvýskyt těchto ustálených eliptických spojení s jevy zvukové povahy (pauza, smích), s navazovacím *no* nebo se střídáním mluvčích:

— *oni tam byli Vojta s NP* ***(smích) ti řikám*** *oni sou kámoši*
— *to právě nevim jesi de.* ***si mysim*** *že jo*
— *tak to všichni viděli* ***no. no. by mě zajímalo*** *kde ho vyhrabali ty jo*
— *a odkdy? </sp><sp>.. počkej mysim že od osumnásteho. </sp><sp>* ***si děláš srandu*** *Jirko. </sp><sp> od osumnásteho srpna no.*

POUŽITÉ KORPUSY

Benešová, L. — Křen, M. — Waclawičová, M. (2013): *ORAL2013: reprezentativní korpus neformální mluvené češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupný z: <http://www.korpus.cz>

Popis dostupný z: <https://wiki.korpus.cz/doku.php/cnk:oral2013>

Pravidla přepisu sond pro korpus ORAL2013 dostupná z: <https://wiki.korpus.cz/lib/exe/fetch.php/cnk:prepisovaci_pravidla_oral2013.pdf>

LITERATURA

Brazil, D. (1995): *A Grammar of Speech*. Oxford: Oxford University Press.

Čermák, F. (1996): Obecná a spisovná čeština: poměr, funkce a metodologie. In: R. Šrámek (ed.), *Spisovnost a nespisovnost dnes*. Brno: Masarykova univerzita, s. 14–18.

Edlund, J. — Heldner, M. — Gustafson, J. (2005): Utterance segmentation and turn-taking in spoken dialogue systems. In: B. Fisseni — H.-C. Schmitz — B. Schröder — P. Wagner (eds.), *Computer Studies in Language and Speech*. Frankfurt am Main: Peter Lang, s. 576–587. Dostupné z: <http://www.speech.kth.se/prod/publications/files/1075.pdf>

Hoffmannová, J. (2001): Syntax institucionálního dialogu. *Naše řeč*, 84, s. 113–120.

Hoffmannová, J. — Kolářová, I. (2013): *Se vám to nelíbí?* Jednoslabičné začátky českých výpovědí / dialogických replik: v běžně mluvené češtině a beletristické stylizaci. *Korpus — gramatika — axiologie*, 07, s. 36–47.

Christiansen, M. H. — Allen, J. (1997): Coping with variation in speech segmentation. In: *Proceedings*. Edinburgh: University of Edinburgh, s. 327–332. Dostupné z: <http://cnl.psych.cornell.edu/pubs/1997-ca-gala.pdf>

Karlík, P. — Nekula, M. — Rusínová, Z. (1995): *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.

Kvale, K. (1993): *Segmentation and Labelling of Speech*. Doctoral dissertation thesis. Trondheim: Norwegian Institute of Technology. Dostupné z: <http://www.diva-portal.org/smash/get/diva2:126345/FULLTEXT01.pdf>

Müllerová, O. (1994): *Mluvený text a jeho syntaktická výstavba.* Praha: Academia.

Müllerová, O. (1996): Výskyt a funkce slova *no* v českých textech prostě sdělovacího stylu. *Stylistyka*, 5, s. 222–229.

Müllerová, O. (2007): Postpozitivní žejo, jo ve vzpomínkovém vyprávění. In: J. Hoffmannová — O. Müllerová (eds.), Čeština v *dialogu generací*. Praha: Academia, s. 291–297.

Passonneau, R. J. — Litman, D. J. (1996): Empirical analysis of three dimensions of spoken discourse: Segmentation, coherence, and linguistic device. In: Hovy, E. H. — Scott, D. R. (eds.), *Computational and Conversational Discourse*. Berlin — Heidelberg: Springer-Verlag, s. 161–194. Dostupné z: <http://www.cs.columbia.edu/~becky/pubs/book93sub.pdf>

Rao, S. — Lane, I. — Schultz, T. (2007): Optimizing sentence segmentation for spoken language translation. In: *Proceedings of Interspeech 2007*. Antverp: ISCA, s. 2845–2848. Dostupné z: <http://www.academia.edu/2847164/Optimizing_sentence_segmentation_for_spoken_language_translation>

Sgall, P. — Hronek, J. (1992): *Čeština bez příkras*. Praha: H+H.

Schwitalla, J. (2012): *Gesprochenes Deutsch*. Berlin: Erich Schmidt Verlag.

Svoboda, A. (2000): Klitika z hlediska funkční větné perspektivy. In: Z. Hladká — P. Karlík (eds.), *Čeština — univerzália a specifika 2*. Brno: Masarykova univerzita, s. 149–159.

Toman, J. (2000): Prosodické spekulace o klitikách v nekanonických pozicích. In: Z. Hladká — P. Karlík (eds.), *Čeština — univerzália a specifika 2*. Brno: Masarykova univerzita, s. 161–166.

Toman, J. (2001): Ertlova diskuse českých klitik. In: Z. Hladká — P. Karlík (eds.), *Čeština — univerzália a specifika 3*. Brno: Masarykova univerzita, s. 73–79.

Jana Hoffmannová | Oddělení stylistiky a lingvistiky textu, Ústav pro jazyk český AV ČR | Letenská 4/123, 118 51 Praha 1
hoffmannova@ujc.cas.cz

Olga Richterová | Ústav Českého národního korpusu, FFUK | nám. Jana Palacha 2, 116 38 Praha 1
olga.richterova@ff.cuni.cz