

Les unités accentuelles asyntagmatiques en français et en tchèque

Tomáš Dubeda

Institut de Phonétique
 Université Charles à Prague, nám. J. Palacha 2, 116 38 Praha 1, République Tchèque
 Tél. : ++420 221 619 250 – Fax : ++420 221 619 382
 Mél : dubeda@ff.cuni.cz – http://fu.ff.cuni.cz

ABSTRACT

The article gives a description of asyntagmatic stress units (stress units corresponding neither to a single word nor to a syntagm) in French and Czech. The structural analysis concerns their lexical occupation, length and frequency. Intonationally, asyntactic stress units show specific preferences as to their tonal configurations, and they are reluctant to initial stress in French. The durational analysis of these units shows different patterns in final lengthening. Globally speaking, asyntactic stress units differ more markedly from single-word-stress-units in French than they do in Czech. The present analysis provides not only a theoretical insight into the given problem, but has an applicative potential also, namely in speech synthesis and recognition.

1. L'UNITÉ ACCENTUELLE EN FRANÇAIS ET EN TCHÈQUE

La fonction segmentative de la prosodie a fait l'objet de nombreuses études, orientées tant vers ses fondements théoriques (p. ex. Boulakia [1], Grosjean et Dommergues [4], Martin [6], Palková [9], Rossi et al. [11]) que vers ses applications, notamment dans la synthèse et dans la reconnaissance de la parole (p. ex. Martin [7], Palková et Ptáček [10]). Il est généralement admis que les liens réunissant la syntaxe et la prosodie sont loin d'être simples et que les processus d'encodage (syntaxe → prosodie) et de décodage (prosodie → syntaxe) diffèrent non seulement dans leur direction, mais aussi dans leur structure.

En tant que constituant prosodique, l'unité accentuelle (connue aussi sous des appellations plus ou moins synonymes de « groupe rythmique » ou de « mot prosodique ») se caractérise par une proéminence (ou plusieurs proéminences, p. ex. dans le cas des unités biaccentuelles en français – Fónagy [3]) ainsi que par une cohérence intérieure (p. ex. le schéma mélodique ou l'allongement progressif des segments vers la fin de l'unité). La table 1 résume, sur un fond contrastif, les traits fondamentaux de l'unité accentuelle (UA) en français et en tchèque (Dubeda [2]) :

Table 1 : Caractéristiques de l'unité accentuelle (UA) en français et en tchèque.

	français	tchèque
position de l'accent	a) finale (sauf les <i>e</i> caducs réalisés) b) initiale (en option ; sauf les syllabes prétoniques)	initiale (sauf les syllabes prétoniques)
caractéristiques duratives	allongement final important	allongement final
caractéristiques intonatives	typiquement, ton bas sur la première syllabe et ton haut sur la dernière syllabe (sauf UA devant frontière prosodique majeure)	
caractéristiques dynamiques	par défaut, déclin d'intensité vers la fin, accompagné dans les unités plus longues d'une montée initiale	
rappports avec la structure lexicale	ancrage fixe par rapport au mot ; accent démarcatif ; la frontière de l'UA implique celle du mot	
longueur lexicale	2,4 mots en moyenne	1,5 mots en moyenne
longueur phonétique	3,25 syllabes en moyenne	

2. LA PHONOSYNTAXE AU NIVEAU DE L'UNITÉ ACCENTUELLE

Contrairement aux approches conventionnelles, surtout générativistes, qui balisent le chemin menant de la syntaxe vers la prosodie, nous situons les composantes prosodiques au départ de notre raisonnement, pour étudier ensuite leurs corrélats syntaxiques. Dans cette perspective, nous proposons une division tripartite des UA (table 2) :

Table 2 : Les trois types syntaxiques des unités accentuelles (« | » – frontière de l'UA).

Type d'UA	Définition	Exemple
unité-mot	UA correspondant à un mot	naturelle ¹ ment
unité-syntagme	UA correspondant à un syntagme	à la mon ¹ tagne
unité asyntagmatique	UA ne correspondant pas à un syntagme ni à un mot	il ¹ veut par ¹ tir

La nature des unités asyntagmatiques prête donc aux conflits entre la structure syntaxique et le découpage prosodique. Il est évident que ce chevauchement est déclenché le plus souvent par des mots grammaticaux (*il* dans l'exemple ci-dessus) qui, ayant le plus souvent une seule syllabe et n'étant pas porteurs d'accent, sont attachés au constituant syntaxique adjacent non pas en congruence avec la syntaxe, mais selon les principes rythmiques. La présence et la structure des unités accentuelles asyntagmatiques sont déterminées par la position de l'accent par rapport à la frontière du mot (situation quasi symétrique en français et en tchèque), par l'ordre des mots (bien plus libre en tchèque qu'en français), par la structure morphosyntaxique (par rapport au français qui est une langue plutôt analytique, les catégories grammaticales sont exprimées en tchèque d'une façon synthétique) et par la longueur des mots (supérieure en tchèque).

3. OBJET DE L'ÉTUDE ET SA MOTIVATION, CORPUS, MÉTHODOLOGIE

L'objet de la présente étude est de donner une description structurelle et prosodique des UA asyntagmatiques, par opposition aux UA correspondant à un mot. Les deux langues que nous mettons en contraste nous permettent d'observer les relations entre la syntaxe et la prosodie dans des conditions différentes (voir le bref portrait du français et du tchèque dans la section précédente). L'analyse phonosyntaxique présentée dans cet article est basée sur un corpus bilingue de lecture oralisée (partie française: locuteur masculin non-professionnel mais cultivé, français suisse non-dialectal, textes journalistiques, 2961 segments, vitesse d'articulation moyenne de 5,48 syll./s; partie tchèque: locuteur masculin semi-professionnel, tchèque standard, textes journalistiques, 3242 segments, vitesse d'articulation moyenne de 5,50 syll./s). L'annotation du corpus a été suivie de l'extraction des paramètres prosodiques (f_0 , durée, intensité). L'étiquetage accentuel, opéré auditivement, a servi de base à la segmentation en unités accentuelles. Par la suite, les UA ont été pourvues d'informations d'ordre syntaxique.

La première partie de la description (section 4) concerne les traits structurels des UA asyntagmatiques et s'étend sur tout le corpus; la seconde partie, prosodique (section 5), se limite aux UA trisyllabiques. L'analyse porte sur la f_0 et sur la durée. Vu les objectifs de la présente recherche, l'évolution des paramètres prosodiques a été stylisée de la façon suivante: a) Pour la f_0 , les données brutes en Hz (la proximité fréquentielle des deux voix étudiées a permis cette simplification), prises sur les noyaux vocaliques des trois syllabes, ont été converties en trois « tons » relatifs: 1 (le plus bas), 2 (intermédiaire) et 3 (le plus haut). A chaque UA correspond donc une des six configurations possibles

{123; 132; 213; 231; 312; 321}. L'analyse intonative ne tient pas compte du contexte. b) La durée des syllabes a été normalisée en fonction de leur structure (nombre de segments, présence de voyelles longues), afin de rendre comparables les syllabes ayant des structures différentes. Par la suite, l'évolution durative des UA a été exprimée en multiples de la première syllabe du groupe; ainsi, la suite « 1-1,11-1,35 » traduit le fait que la deuxième syllabe de l'UA (après normalisation) dépasse de 11% la première et que la dernière syllabe de l'UA est de 35% plus longue que la première.

4. LES UA ASYNTAGMATIQUES – ASPECTS STRUCTURELS

L'analyse du corpus montre que la part des UA asyntagmatiques sur le total des UA est de 33% en français et de 27% en tchèque. Le fait que ce pourcentage est plus élevé en français est certainement dû à la grammaire analytique de cette langue, qui prête à des découpages asyntaxiques. Ceci se traduit également dans la composition des UA asyntagmatiques: celles-ci peuvent contenir jusqu'à 7 mots en français, contre un maximum de 4 mots en tchèque. La longueur moyenne de l'UA asyntagmatique est de 3,91 syllabes (correspondant à 2,71 mots) en français et de 3,94 syllabes (correspondant à 2,30 mots) en tchèque.

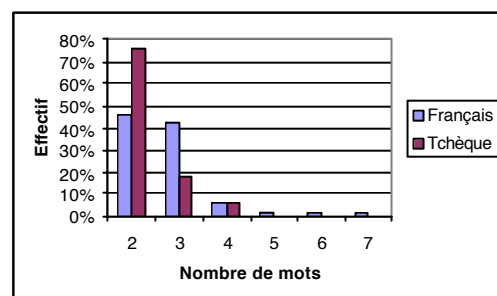


Figure 1 : Répartition des UA asyntagmatiques en fonction du nombre de mots qu'elles contiennent.

Une autre caractéristique importante est la structure syntaxique des UA asyntagmatiques. Le tableau suivant donne un aperçu de la composition syntaxique des UA asyntagmatiques à deux mots:

Table 3 : Occupation syntaxique des UA asyntagmatiques à deux mots (MSP – mot à sens plein).

tchèque		français	
deux MSP	11%	deux MSP	2%
préposition + MSP	14%		
MSP + préposition	11%	préposition + MSP	40%
MSP + conjonction	9%	conjonction + MSP	7%
MSP + auxiliaire	8%	auxiliaire + MSP	2%
MSP + adverbe	14%	adverbe + MSP	4%
MSP + déterminant	6%	déterminant + MSP	33%
MSP + pronom	15%	pronom + MSP	7%
autres	14%	autres	4%

L'analyse structurale des UA asyntaxiques montre les tendances suivantes :

a) Abstraction faite des UA contenant deux mots à sens plein, l'UA asyntaxique se compose généralement d'un mot à sens plein et d'un mot grammatical ; l'ordre de ces deux composantes est inverse dans les deux langues, ce qui s'explique par la position de l'accent.

b) De là résulte aussi la différence dans le branchement syntaxique des mots à l'intérieur des UA asyntaxiques : en français, le mot grammatical au début de l'UA est immédiatement rattaché à un syntagme couvrant deux UA, p. ex. *des cen'taines* | *de poi'ssons* (le signe «|» dénote la frontière de l'UA). Ceci n'est pas le cas en tchèque, où le mot grammatical à la fin de l'UA est immédiatement rattaché au mot suivant, p. ex. *'vzdálen své* | *'vlasti* « éloigné de sa partie ».

c) Étant donné que les prépositions monosyllabiques en tchèque reprennent très souvent l'accent du mot qui suit, nous trouvons dans cette langue aussi bien des UA « préposition + MSP » que des UA « MSP + préposition ». En français, il n'existe que le type « préposition + MSP ».

5. LES UA ASYNTAGMATIQUES – ASPECTS PROSODIQUES

Afin d'opérer, sur une base contrastive, l'analyse prosodique des UA asyntaxiques, nous n'avons retenu que les UA trisyllabiques, très fréquentes dans les deux langues et présentant assez d'espace pour l'étude des configurations intonatives et duratives. Par souci d'homogénéité de l'échantillon, nous avons exclu les UA prépausales et conclusives.

5.1 Fréquence fondamentale

Les figures 2 et 3 apportent des renseignements sur la structure intonative des UA asyntaxiques par comparaison à celles qui contiennent un seul mot :

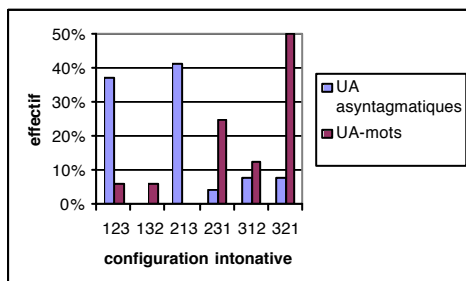


Figure 2 : Structure intonative des UA asyntaxiques et des UA-mots en français ; pour la lecture des configurations intonatives, voir section 3.

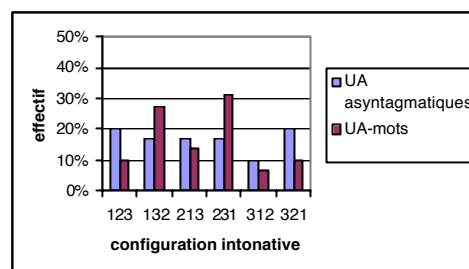


Figure 3 : Structure intonative des UA asyntaxiques et des UA-mots en tchèque ; pour la lecture des configurations intonatives, voir section 3.

Les données présentées dans les graphiques se résument dans les tendances suivantes :

a) En français, la structure syntaxique des UA se traduit assez nettement dans les contours intonatifs : les UA-mots sont réalisées pour la plupart avec une descente finale, tandis que les UA asyntaxiques sont corrélées avec le dernier ton montant. La différence est certainement due à l'accent initial qui est assez fréquent dans le corpus et qui correspond le plus souvent au premier ton 1 ou 2 : il frappe 37% des UA-mots, contre 4% des UA asyntaxiques.

b) En tchèque, les différences entre les deux types d'UA sont moins marquées. De manière générale, les UA asyntaxiques présentent les profils 123 et 321 (unidirectionnels, sans inflexion intonative) plus souvent que les UA-mots et leur syllabe intermédiaire correspond au maximum intonatif bien moins souvent que dans le cas des UA-mots.

c) Quant à l'envergure intonative (différence en Hz entre les tons 1 et 3), nous constatons que celle-ci est généralement plus importante dans les unités asyntaxiques (32 Hz en français, 4,6 Hz en tchèque) que dans les unités-mots (22,8 Hz en français, 4,34 Hz en tchèque). L'envergure intonative est un facteur différenciateur qui oppose le tchèque au français (Materna [8]).

5.2 Durée

Les figures 3 et 4 expriment les configurations duratives des UA étudiées :

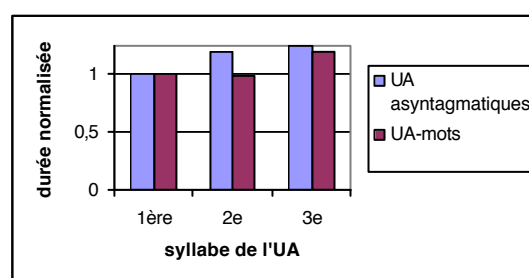


Figure 3 : Structure durative des UA asyntaxiques et des UA-mots en français.

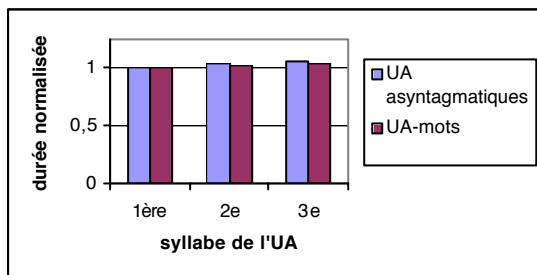


Figure 4 : Structure durative des UA asyntagmatiques et des UA-mots en tchèque.

Dans les deux langues, nous constatons, conformément aux recherches antérieures, un allongement progressif des syllabes à l'intérieur des UA, à l'unique exception des UA-mots en français, qui peut être expliquée d'une part par la présence de l'accent initial (la composante durative de ce type d'accent, quoique faible, a été confirmée p. ex. par Jankowski et al. [5]), et d'autre part par l'abrègement des mots grammaticaux, figurant le plus souvent au début de l'UA. De manière générale, l'évolution durative des syllabes est moins marquée en tchèque, mais les deux langues convergent sur un point : l'allongement est plus marqué dans les UA asyntagmatiques.

6. CONCLUSION ET PERSPECTIVES

L'étude descriptive d'un corpus bilingue de parole lue a mis en évidence quelques aspects phonosyntaxiques des unités accentuelles asyntagmatiques, qui se distinguent par une plus grande longueur syllabique par rapport aux unités-mots et dont la fréquence et l'occupation lexicale sont nettement plus grandes en français qu'en tchèque. Vu les propriétés grammaticales des deux langues, les UA asyntagmatiques commencent typiquement par un mot grammatical en français, et finissent par celui-ci en tchèque. En outre, les UA asyntagmatiques inhibent la mise en place de l'accent initial en français, ce qui se répercute dans leur comportement intonatif. Les deux types d'UA montrent des préférences spécifiques quant aux configurations tonales, et, dans une plus faible mesure, quant à l'envergure intonative. La différenciation durative des syllabes au sein des UA-mots et des UA asyntagmatiques semble être plus forte en français.

En dehors des observations phonétiques et phonologiques, la présente enquête peut être soumise à une formalisation (qu'elle soit déterministe ou stochastique) visant à son application dans le module syntaxico-prosodique d'un système de synthèse ou de reconnaissance de la parole.

Conscient des limites de la présente analyse (notamment au niveau de leur valeur statistique), nous comptons, dans un proche avenir, inclure dans le cadre descriptif d'abord des UA asyntagmatiques de différentes longueurs, pour ensuite compléter l'analyse

prosodique du troisième paramètre – l'intensité. Une description tenant compte du contexte prosodique (notamment dans le cas de l'intonation) apporterait également des résultats plus riches.

REMERCIEMENTS

La publication de cette communication a été rendue possible grâce au projet GA AV B9101402. L'auteur remercie l'équipe du LAIP (Université de Lausanne) pour les outils de segmentation automatique de la parole, ainsi que les deux locuteurs.

BIBLIOGRAPHIE

- [1] G. Boulakia. *Phonosyntaxe*, thèse de doctorat, Université Paris 7, 1974.
- [2] T. Dubeda. Structural and quantitative properties of stress units in Czech and French. In *Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*. A. Braun and H. R. Masthoff (eds.), Stuttgart : Steiner, 2002.
- [3] I. Fónagy. L'accent français : accent probabilitaire (Dynamique d'un changement prosodique). In *L'accent en français contemporain*, Didier, Ottawa, 1979.
- [4] F. Grosjean and J.-Y. Dommergues. Les structures de performance en psycholinguistique. *L'année psychologique*, 83, pages 513–536, 1983.
- [5] L. Jankowski, C. Astésano and A. Di Cristo. The initial rhythmic accent in French : Acoustic data and perceptual investigation. In *Proceedings of the 14th Congress of Phonetic Sciences, San Francisco*, pp. 257–260, 1999.
- [6] P. Martin. Sur les principes d'une théorie syntaxique de l'intonation. In : *Problèmes de prosodie*, P. Léon and M. Rossi (eds.), Didier, Ottawa, vol. I, 1979.
- [7] P. Martin. Association prosodie-syntaxe : validation par synthèse. In : *Actes des XXII^{èmes} Journées d'Etudes sur la Parole*, Martigny, pp. 119-122, 1998
- [8] P. Materna. *Mémoire et mélodie : Le français est-il plus mélodieux que le tchèque?* Mémoire de maîtrise, Université Charles à Prague, 2002.
- [9] Z. Palková. Discourse segmentation and phrase structure. In *Acta Universitatis Carolinae, Phonetica Pragensia VI*, Univerzita Karlova, 1980.
- [10] Z. Palková and M. Ptáček. Modelling Prosody in TTS Diphone Synthesis in Czech. In *Speech processing, Forum Phonetikum 63*, Hector, Frankfurt am Main, pages. 59–77, 1997.
- [11] M. Rossi, A. Di Cristo, D. Hirst, P. Martin and Y. Nishinuma. *L'intonation : de l'acoustique à la sémantique*, Klincksieck, 1981.