

## La surabondance des formes dans la flexion croate

Tomislava Bošnjak Botica  
Institut de la langue croate et de linguistique\*

Sara Košutar  
L'université de Tromsø – L'université arctique de Norvège\*\*

---

La surabondance des formes dans la flexion se réfère à la présence de multiples formes pour exprimer le même sens grammatical. Depuis peu, ce phénomène a suscité un grand intérêt dans la recherche linguistique, notamment dans les langues qui étaient moins investiguées. Dans cet article, on examine de plus près la surabondance morphologique en croate, en mettant l'accent sur les raisons possibles de cette variation et ses implications pour la description grammaticale du croate mais aussi d'autres langues proches. On présente la *Base de données des doublets croates* qui peut servir comme un outil bienvenu pour une connaissance plus approfondie de ce phénomène.

*Mots clés* : surabondance morphologique, Base de données doublets, flexion, langue croate.

Dans certains courants de la linguistique du XXe siècle, principalement provenant du structuralisme et de la grammaire générative, une opinion largement répandue soutenait que chaque forme linguistique correspondait à un seul sens, c'est-à-dire qu'une forme était associée à une fonction grammaticale spécifique (et vice versa). Ce phénomène est appelé le principe de blocage morphologique (angl. *Morphological Blocking Principle*, v. Aronoff 1976) ou le principe de contraste (angl. *Principle of Contrast*, cf. Clark 1987 ; Carstairs-McCarthy 2010). Selon ces principes, l'existence d'une forme empêche automatiquement

---

\* tbošnjak@ihjj.hr.

\*\* sara.kosutar@uit.no.

Ce travail a été soutenu par la Fondation croate des sciences dans le projet intitulé *Multilevel approach to spoken discourse in language development* (UIP-2017-05-6603) et par le Conseil de recherche en arts et sciences humaines dans le projet intitulé *Feast and Famine: Confronting overabundance and defectivity in langue* (AH/T002859/1).

l'apparition d'une autre forme ayant le même sens (ou la même fonction). Il s'agit d'un principe économique qui est propre à la langue. Par exemple, en croate, la forme supplétive *ljudi*-NOM.PL.M 'hommes/gens' du mot *čovjek*-NOM.SG.M 'homme' bloque l'apparition de la forme \**čovjeci*, qui est faite suivant le modèle de la déclinaison attendue. Dans les langues indo-européennes, la conjugaison du verbe *être* comporte des formes supplétives à des temps différents, comme en français (*je suis* ~ *j'étais* ~ *je fus*) ou en anglais (p. ex., *I am*-PRS.1.SG 'je suis' ~ *I was*-PST.1.SG 'j'étais' / 'je fus'). Les adjectifs au comparatif présentent ce phénomène aussi, comme la formation du comparatif en français *bon* > *meilleur* ou en anglais *good* > *better*.

En croate, la grande majorité des lexèmes suivent le schéma d'une forme, par exemple *konj*-NOM.SG.M 'cheval' réalise le pluriel *konji*-NOM.PL.M et non \**konjevi*, alors que *ovan*-NOM.SG.M 'bélier' devient *ovnovi*-NOM.PL.M et non \**ovni* au pluriel. Il existe également des lexèmes qui peuvent prendre deux formes, mais il semble que ces formes doubles soient généralement associées à une différence fonctionnelle. Par exemple, *sat*-NOM.SG.M 'heure' devient *sati*-NOM.PL.M 'heures' au pluriel pour désigner des tranches de temps de 60 minutes, tandis que le pluriel *satovi*-NOM.PL.M désigne les dispositifs pour mesurer le temps. Donc, la différence dans la forme reflète la différence dans le sens (en français, il s'agit de deux lexèmes différents). Ce type de variation peut être dû à des facteurs historiques, régionaux, sociolinguistiques, stylistiques, etc.

Cependant, nous savons par notre expérience linguistique et les descriptions linguistiques disponibles qu'il existe également des formes doubles, c'est-à-dire que pour le même sens grammatical, il peut y avoir parfois deux ou plusieurs formes qui ne sont pas forcément liées à une différence fonctionnelle. On utilise ici le terme *double*, *doublet* ou *doubletisme*, mais le phénomène se réfère à trois (ou plusieurs) formes également. Ceci peut être illustré par les exemples en croate: le mot *kuhar*-NOM.SG.M 'cuisinier/chef' peut avoir deux formes à l'instrumental singulier *kuharom/kuharem*; le verbe *šetati se*-INF 'se promener' peut avoir deux formes au présent *šetam/šećem se*-PRS.1.SG; l'adjectif *dug*-NOM.SG.M 'long' a deux formes au comparatif *duži/dulji*, le nom *naranča*-NOM.SG.F 'orange' au génitif pluriel peut être réalisé comme *naranča*, *naranči* ou *naranača* etc. Ces formes qui soit proviennent d'une même base mais ont des morphèmes flexionnels différents, soit subissent une allomorphie dans la base, apparaissent dans les mêmes contextes grammaticaux et sémantiques. Par conséquent, il n'y a pas de différence fonctionnelle nette entre ces deux (ou trois) formes.

### 1.1. La notion de la surabondance

Le phénomène des formes doubles différentes mais avec le même sens grammatical et lexical est appelé *surabondance* (voir Thornton 2011, 2012). La

surabondance morphologique d'un côté et la défektivité de l'autre, sont actuellement un domaine de recherche très fructueux en linguistique. Le phénomène de la surabondance lui-même a été développé à partir de la typologie canonique de G. G. Corbett (2007, 2009). Corbett (2009 : 1) définit le système flexionnel canonique comme un système dans lequel il existe une correspondance unique entre la forme et la fonction, ainsi qu'entre la fonction et la forme. Cela signifie qu'une forme donnée est associée à une fonction grammaticale spécifique et vice versa. Cependant, la surabondance représente un phénomène non canonique, c'est-à-dire une déviation de la canonicité, car deux ou plusieurs formes peuvent occuper la même place (cellule) dans le paradigme sans que l'utilisation de l'une ou l'autre forme soit conditionnée par un contexte situationnel spécifique. Cela signifie que les locuteurs peuvent choisir librement entre ces formes sans qu'il y ait de contrainte contextuelle.

Dans certaines langues, ce phénomène est rarement représenté. Par exemple, en anglais, les formes doubles ne se manifestent que dans le pluriel de plusieurs noms d'origine latine (p. ex., le mot *formula* 'formule' > *formulas/formulae* 'formules') ou dans le participe passé de certains verbes (p. ex., *leap* 'sauter' > *leapt/leaped* 'sauté'). Dans les langues romanes, on trouve également quelques exemples isolés, tels que le participe passé de certains verbes en italien (p. ex., *seppellire* 'enterrer' > *sepolto/seppellito* 'enterré') ou dans la langue française, la 1<sup>ère</sup> personne du singulier du verbe *pouvoir* (*je peux/puis*) et les formes des verbes *s'asseoir* et *balayer* au présent, à la fois au singulier et au pluriel (p. ex. *je m'assois/assieds*-PRS.1.SG, *ils balaient/balayent*-PRS.3.PL). En revanche, les langues slaves offrent un terrain très fertile pour étudier ce phénomène, car on y trouve de nombreuses occurrences de formes doubles. Ces langues possèdent un système de flexion extrêmement riche, et les doublets morphologiques englobent souvent des catégories entières de lexèmes, et non seulement des exemples isolés. Ainsi, en ce qui concerne le croate (v. Bošnjak Botica & Hržica 2016 ; Bošnjak Botica *et al.* 2018 ; Bošnjak Botica *et al.* 2019 ; Jelaska & Bošnjak Botica 2019 ; Lečić 2015), les formes doubles se rencontrent dans la déclinaison des noms (p. ex., *ribarom/ribarem*-INSTR.SG.M 'pêcheur' *zubi/zuba/zubiju*-GEN.PL.M 'dents' ; vocative singulier *izdajico/izdajice* 'traître') et des pronoms (p. ex., *mojeg/mog*-GEN.SG.M 'mon', *njegovu/njegovom*-DAT.SG.M 'son'), la conjugaison des verbes (p. ex., *šetam/šećem se*-PRS.1.SG 'je me promène' ; *maći/maknuti*-INF 'enlever' ; *donesen/donijet*-PTCP.SG.M 'amené'), la déclinaison des adjectifs (p. ex., *bratova/bratovog*-DAT.SG.M 'de frère'), la gradation des adjectifs (p. ex., *čistiji/čišći*-NOM.SG.M 'plus propre') ainsi que des adverbes (p. ex., *mrskije/mrže* 'plus odieusement', et la formation de certaines catégories de noms (p. ex., *estet/esteta*-NOM.SG.M 'esthète').

La surabondance morphologique a longtemps été négligée dans la recherche linguistique pour diverses raisons (v. Bošnjak Botica *et al.* 2019). Tout d'abord, elle ne correspond pas à la description dans des grammaires tradition-

nelles (plutôt normatives) qui postulent une relation univoque entre la forme et le sens, considérant la variabilité formelle comme indésirable. Deuxièmement, bien qu'étant considérablement attestée dans certaines langues, la surabondance morphologique est un phénomène rare, limité à un nombre restreint de paradigmes et de catégories grammaticales, même si dans certaines langues ils y existent. Ces facteurs ont sans doute contribué à un manque de recherche approfondie sur le sujet.

Les raisons de cette négligence peuvent également être d'ordre méthodologique. En raison de sa nature quand même marginale dans la langue, la surabondance morphologique est souvent difficile à percevoir et à définir avec précision. De plus, dans certains cas, il est difficile d'établir une distinction claire entre les formes standard et non standard, car les manuels normatifs peuvent différer dans leurs opinions, et ces distinctions ne sont pas toujours documentées de manière convaincante. Les dictionnaires de la langue moderne fournissent des données variées et parfois moins fiables concernant la surabondance.

Cependant, ces obstacles ont commencé à être surmontés récemment, et la surabondance morphologique suscite désormais un intérêt croissant dans la recherche linguistique. Des études plus approfondies sont menées pour comprendre les mécanismes et les implications de ce phénomène (p. ex., Bermel & Knittl 2012 ; Bermel *et al.* 2015 ; Cappellaro 2013 ; Cummins 1995 ; Fehringer 2004 ; Mörth & Dressler 2014 ; Naghzhuy-Kohan & Kuteva 2016 ; Santilli 2014 ; Thornton 2011). Les avancées méthodologiques et l'accès à des corpus linguistiques plus vastes ont également contribué à une meilleure compréhension de ce phénomène, qui à son tour permet une meilleure description de la complexité et de la diversité des systèmes linguistiques, ainsi que des variations qui existent au sein des langues elles-mêmes.

### *1.2. La base des données Dvojba est la recherche des formes doubles en croate*

La recherche en linguistique contemporaine s'appuie de plus en plus sur les technologies modernes, notamment dans le domaine de la morphologie, où la description morphologique des langues est souvent réalisée à l'aide d'outils technologiques (p.ex. dans la recherche de corpus). Cependant, dû à la nature marginale de la surabondance morphologique, les outils linguistiques disponibles peuvent souvent traiter de tels exemples de manière incorrecte ou incomplète, ce qui peut compromettre la fiabilité des résultats obtenus.

Les sources linguistiques traditionnelles (les dictionnaires, les grammaires, même les corpus) utilisées dans la recherche linguistique ne sont pas toujours adaptées pour récupérer facilement ces données spécifiques (v. Bošnjak Botica *et al.* 2018). C'est pourquoi il serait bénéfique de développer des outils plus ciblés

qui pourraient aider les linguistes et d'autres utilisateurs de la langue à vérifier le statut de certaines formes. Ces outils spécialisés pourraient permettre une analyse plus précise des formes doubles et de la surabondance morphologique en général. Ils pourraient aider à identifier et à catégoriser les exemples des formes doubles dans différentes langues, ainsi qu'à établir des distinctions entre les formes standard et non standard. De tels outils faciliteraient la collecte de données fiables et complètes sur la surabondance morphologique, ce qui contribuerait à approfondir notre compréhension de ce phénomène linguistique.

*La base de données des doublets morphologiques croates (Dvojba, Bošnjak Botica et al. 2019 ; voir [www.dublete.jezik.hr](http://www.dublete.jezik.hr)<sup>1</sup>)* constitue une ressource linguistique accessible au public et facilement consultable. Son principal objectif est de soutenir la recherche scientifique sur ce phénomène en fournissant un point de référence fiable pour l'analyse des unités et catégories spécifiques de doublets morphologiques. Cette base de données vise à répertorier et à regrouper un nombre représentatif de formes doubles issues des corpus existants de la langue croate, ce qui permet d'explorer et de décrire ce phénomène et de comprendre sa place dans la structure du système linguistique.

Les données utilisées dans cette base (la forme double, contexte, lemme, auteur, l'année etc.) ont été extraites de deux corpus linguistiques croates: *Hrvatska jezična riznica* (Riznica, Čavar & Brozović Rončević 2012) et *Hrvatski mrežni korpus (hrWaC ; Ljubešić & Klubička 2014)* à l'intermédiaire d'une table Excel, qui est alors importée dans la Base. Les données ont été vérifiées manuellement, si nécessaire. Une des principales différences entre les deux corpus réside dans le caractère contemporain de hrWaC contrairement à Riznica qui comporte des textes dont la création est beaucoup plus étalée dans le temps.

Ces corpus représentent une vaste collection de textes en croate provenant de différentes sources, offrant ainsi une diversité de contextes linguistiques dans lesquels les formes doubles peuvent apparaître. En recueillant et en analysant ces données, la base de données des doublets morphologiques croates contribue à la recherche et à la compréhension de ce phénomène particulier dans la langue croate.

À notre connaissance, il n'existe pas de base de données des doublets morphologiques dans d'autres langues, ce qui rend cette ressource linguistique croate d'autant plus innovante. Grâce à cette ressource, les chercheurs disposent d'un outil bienvenu pour approfondir leur compréhension de la surabondance morphologique et pour mener des analyses comparatives entre différentes langues.

La base de données des doublets morphologiques contient actuellement environ 61 800 entrées, ce qui en fait une ressource linguistique considérable et riche en informations. Chaque entrée de la base de données correspond à l'attes-

---

<sup>1</sup> Dernier accès : 15/02/2023.

tation d'une forme double spécifique dans le corpus. Pour chaque entrée, plusieurs valeurs ont été déterminées afin de fournir des détails précis sur les caractéristiques morphologiques et contextuelles de la forme. Les catégories morphologiques de chaque entrée ont été codées selon la Norme MULTEXT-Est (Erjavec *et al.* 2003), qui offre une standardisation permettant une analyse comparative plus facile des données. Ces catégories comprennent des informations telles que la classe de mots, le genre, le nombre, le cas, la personne, etc.

En plus des catégories morphologiques, chaque entrée de la base de données comprend également des informations sur l'œuvre dans laquelle la forme apparaît, l'auteur de l'œuvre, l'aire géographique de l'auteur, le temps de création et la source (corpus) où la forme a été trouvée. Ces informations contextuelles sont essentielles pour une analyse approfondie des formes doubles et pour situer leur utilisation dans un contexte spécifique. Par exemple, il est possible de rechercher des formes doubles spécifiques selon des critères tels que l'auteur, où l'on peut voir quel auteur préfère quelle forme ou s'il utilise les deux formes différentes (doubles) dans son œuvre, et pour quelles raisons. On peut aussi voir l'usage des formes selon différentes périodes ou bien suivre une forme dès ses premières attestations (si elles sont disponibles) jusqu'à aujourd'hui pour vérifier son usage.

La base *Dvojba* offre aux utilisateurs la possibilité de répondre à diverses questions liées à l'utilisation et à la distribution des formes doubles. Voici quelques exemples de questions auxquelles les utilisateurs pourraient trouver des réponses en utilisant cette ressource.

*Certaines formes disparaissent-elles et à quelle vitesse ?* Vu qu'elle dispose des données recueillies des corpus ayant une couverture temporelle différente (*Riznica* contient aussi des œuvres du XVIII<sup>e</sup> et XIX<sup>e</sup> siècle et même plusieurs œuvres plus vieux, la base de données permet de suivre l'évolution des formes doubles dans le temps et d'observer si certaines formes sont en déclin ou elles ont disparu complètement. En examinant les informations sur l'année de création des œuvres où les formes apparaissent, les utilisateurs peuvent évaluer la persistance ou la disparition de certaines formes au fil du temps.

*Quelles formes apparaissent le plus souvent ?* La fréquence d'apparition des formes doubles dans la base de données, permet aux utilisateurs de déterminer quelles formes sont les plus stables ou les plus préférentielles dans la langue. Par exemple, à quoi ressemble la répartition des trois suffixes au génitif pluriel des noms féminins comme *naranča* 'orange' ? En étudiant les occurrences des trois suffixes possibles au génitif pluriel des noms féminins, on pourrait observer la répartition des formes et déterminer s'il existe des préférences ou des schémas particuliers dans la répartition des trois suffixes au génitif pluriel. Cela permet de comprendre quel suffixe est le plus fréquemment utilisé et s'il existe des variations en fonction du lexème ou d'autres facteurs.

*La répartition des formes concurrentes dépend-elle de la période ?* En analysant les données de la base en fonction de la période de création des œuvres, les utilisateurs pourraient évaluer s'il existe des variations temporelles dans la répartition des formes doubles. Ils pourraient identifier des tendances ou des changements dans l'utilisation des différentes formes au fil du temps.

*Des tendances différentes émergent-elles en fonction du sens ?* En examinant les formes doubles associées à des sens spécifiques (à partir du contexte de l'emploi), les utilisateurs pourraient déterminer s'il existe des préférences ou des tendances particulières en ce qui concerne l'utilisation des formes en fonction du sens. Cela permettrait de mieux comprendre les motivations derrière l'utilisation de certaines formes et voir si certaines formes sont réservées à certains contextes (grammaticaux ou sémantiques).

L'aspect diachronique de la base de données des doublets morphologiques est essentiel pour comprendre l'évolution et la distribution des formes doubles dans la langue croate. En suivant le moment d'entrée des formes dans la langue et en examinant leur évolution au fil du temps, on peut identifier les changements qui ont eu lieu et les tendances qui se sont développées. On peut, dans une certaine mesure, observer si certaines formes doubles ont disparu de l'usage courant et à quelle vitesse cela s'est produit. Cela permettrait de déterminer si les formes doubles sont une caractéristique temporaire (transitoire) ou si elles ont une durée de vie plus longue (même permanente) dans la langue. Parallèlement, les données synchroniques de la base fournissent un aperçu de l'état actuel des formes doubles. On peut analyser la fréquence d'utilisation des différentes formes, identifier les formes les plus couramment utilisées et examiner s'il y a des variations en fonction du lexème, de la période ou d'autres facteurs linguistiques.

L'aspect diatopique de la base de données des doublets morphologiques est en effet crucial pour comprendre la diffusion et la distribution des formes doubles dans différents lieux géographiques. En examinant les données diatopiques, on peut chercher à déterminer le lieu d'entrée des formes doubles dans la langue croate, ainsi que la direction de leur diffusion.

En comparant les données diachroniques et diatopiques, on peut également répondre à la question de savoir si la surabondance des formes doubles est un phénomène intralinguistique, c'est-à-dire qu'il se produit naturellement dans le système linguistique de la langue croate, ou s'il est d'une certaine manière influencée par le contact avec d'autres langues ou systèmes linguistiques proches.

De même, en analysant les données diachroniques et diatopiques de la base de données des doublets morphologiques, il est possible d'explorer des questions plus approfondies sur le développement et l'évolution des formes doubles dans la langue croate.

En examinant le développement d'une forme double particulière du point de vue diachronique, en tenant compte de son lieu et de son moment d'entrée



dans la langue, on peut obtenir des informations précieuses sur les tendances et les directions du développement futur de la synonymie morphologique dans la langue croate. Par exemple, on pourrait observer si une forme prédomine à certaines époques ou dans certaines régions, ou si de nouvelles formes doubles émergent au fil du temps.

De plus, en étudiant les aires géographiques des auteurs associées aux formes doubles, on peut retracer l'origine de la surabondance en tant que phénomène intralinguistique ou en tant que conséquence des contacts linguistiques avec d'autres systèmes linguistiques. Par exemple, si l'on observe une concentration de formes doubles dans les régions frontalières ou dans les zones où des langues voisines sont parlées, cela pourrait indiquer une influence externe sur la surabondance morphologique. Pour ces raisons, un axe contenant des données dialectales sera aussi inclus.

En outre, on espère pouvoir contribuer au débat actuel sur la nature de la langue en examinant le rôle des formes doubles. Nous pourrions étudier si la langue est un système « parfait », dans lequel les formes doubles sont considérées comme des phénomènes éphémères dans le processus de transition d'une forme à une autre, ou si la langue est plutôt un système plus lâchement organisé, caractérisé par une certaine souplesse et une tolérance aux variations internes.

Enfin, nous pourrions comparer les données des formes doubles avec d'autres notions linguistiques, telles que la productivité morphologique, la régularité des paradigmes et la stabilité des formes. Cela nous aiderait à évaluer si les formes doubles sont des exceptions occasionnelles ou si elles reflètent des tendances plus générales dans l'organisation du langage.

## *2. Les exemples de la recherche des catégories surabondantes en croate*

Les exemples de la recherche présentés ici font partie du projet international *Feast and Famine: Confronting Overabundance and Defectivity in Language* (AH/T002859/1), qui vise à étudier les deux phénomènes marginaux mais fascinants — la surabondance, c'est-à-dire les formes multiples d'un côté, et l'absence totale de forme, de l'autre. Le projet aborde ces questions à partir de différentes perspectives : *Comment ces phénomènes sont-ils décrits dans les manuels ? Que nous montre le corpus ? Comment les locuteurs (adultes et enfants) traitent ce phénomène en situation de choix ?* En outre, il cherche à développer des modèles informatiques spéciaux pour prédire ces phénomènes. Ainsi, la base *Dvojba* est un outil précieux qui peut aider à répondre aux questions soulevées dans le cadre du projet mentionné.



La surabondance morphologique en croate se situe dans quelques 40 catégories qui se rapportent aux noms, pronoms, adjectifs (et adverbes) et verbes (Bošnjak Botica & Hržica 2016 ; Bošnjak Botica *et al.* 2018 ; Bošnjak Botica *et al.* 2019). Certaines catégories sont plus nombreuses, avec un millier de membres (p. ex. les substantifs féminins qui se terminent en *-ost* ayant deux formes à l'instrumental singulier, comme *mladost* 'jeunesse'), alors que d'autres ne contiennent qu'un seul lexème. Une catégorie peut se rapporter à un seul cas ou une seule personne dans le paradigme. Dans la base *Dvojba*, toutes ces catégories sont identifiées, mais elles ne sont pas encore complètes. Dans des cas où une catégorie peut contenir plus d'un millier de lexèmes, les plus fréquents ont été incorporés dans la base.

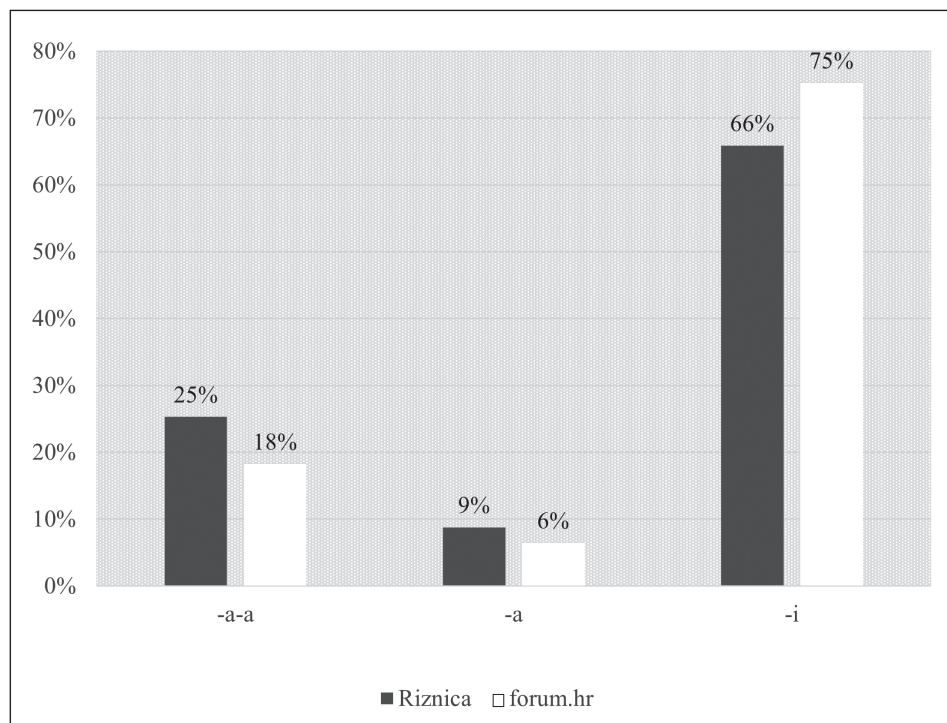
**2.1.** Une de plus nombreuses catégories où la surabondance des formes se réalise est celle du génitif pluriel des noms féminins ayant le groupe consonantique devant le morphème flexionnel. Cette catégorie aborde un phénomène rare dans la langue: la réalisation de trois formes, c'est à dire, trois désinences (*-a*, *-a-a*, *-i*) dans le même cas du paradigme (p. ex., *kruška*-NOM.SG.F 'poire' : *krušaka/kruška/kruški*-GEN.PL.F ; *primjedba*-NOM.SG.F 'remarque' : *primjedaba/primjedba/primjedbi*-GEN.PL.F ; *trešnja*-NOM.SG.F 'cerise' : *trešanja/trešnja/trešnji*-GEN.PL.F). Une partie de ces noms quand-même ne se réalise qu'en deux formes (p. ex., *suradnja*-NOM.SG.F 'coopération' : *suradnja/suradnji*-GEN.PL.F, *bomba*-NOM.SG.F 'bombe' : *bomba/bombi*-GEN.PL.F). Figure 1 montre la surabondance des formes génitives plurielles par rapport au nominatif singulier des noms, respectivement *trešnja* 'cerise', *kruška* 'poire', *lignja* 'calmar', *školjka* 'coquille', *izložba* 'exposition', *kocka* 'cube', *suradnja* 'collaboration', *bomba* 'bombe', *varijanta* 'variante' et *crkva* 'église'.

**Figure 1.** La surabondance des formes génitives plurielles par rapport aux formes au nominatif singulier



Dans Figure 2 (v. Bošnjak Botica *et al.* 2022a), on peut observer les proportions des trois formes dans le corpus de langue plutôt littéraire ou formelle (*Riznica*, Čavar & Brozović Rončević 2012) et dans le corpus de langage non-littéraire, qui contient des échantillons de langue parlée (le sous-corpus *forum.hr* du *hrWaC*, Ljubešić & Klubička 2014). Le calcul présente la valeur moyenne pour les trois désinences des 239 noms ayant au moins 20 attestations au génitif pluriel dans les deux corpus.

**Figure 2.** Les proportions des trois formes au génitif pluriel féminin dans les corpus *Riznica* et *forum.hr*



Les formes en *-i* sont prédominantes à la fois dans le corpus de langue moins formelle ou non-littéraire (*forum.hr*, Ljubešić & Klubička 2014) et dans le corpus de langue plus formelle ou littéraire (*Riznica*, Čavar & Brozović Rončević 2012). La distribution de la forme en *-a* est la moins représentée dans les deux corpus. Ce sont des résultats préliminaires à partir desquelles notre recherche se poursuit pour examiner d'autres questions importantes, comme celles de la comparaison des formes citées dans les entrées du dictionnaire avec le corpus du croate standard contemporain pour examiner les différences entre les formes du dictionnaire et l'usage actuel de la langue standard, aussi bien que les différences entre deux

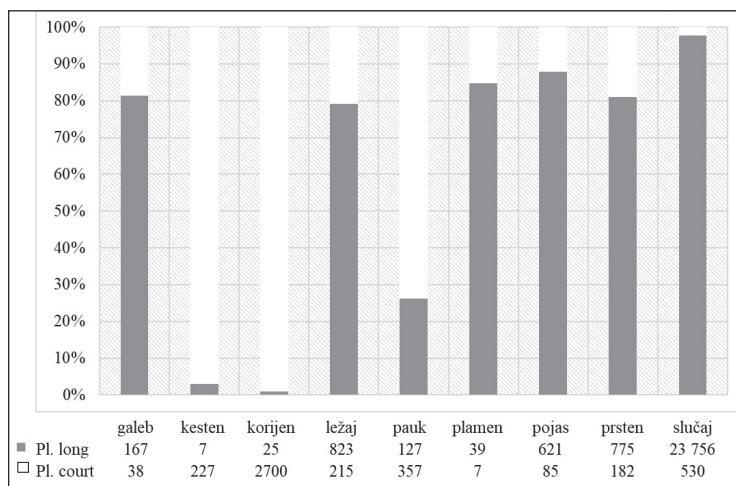
corpus, ce qui nous permettra d'approximer dans quelle mesure la terminaison *-i* s'est répandue au-delà de l'usage autorisé dans la langue standard contemporaine. Nos premiers résultats suggèrent qu'avec la plupart des noms, la surabondance dans le génitif pluriel pourrait rapidement diminuer en faveur de la désinence *-i*. Un des premiers pas à suivre est de bien vérifier les autres caractéristiques qui peut-être ont pu influencer la distribution des formes (p. ex., nombre de syllabes, groupe consonantique, type d'accent, l'origine des mots, la valeur diachronique, l'âge des locuteurs, etc.).

**2.2.** Le deuxième cas qu'on présente ici concerne la surabondance dans la formation du pluriel des noms masculins bisyllabiques se terminant par une consonne. En croate, en principe, les noms masculins forment leur pluriel en fonction du nombre de syllabes. Les noms avec plus d'une syllabe prennent « seulement » les désinences du pluriel (NOM-*i*, GEN-*a*, DAT/LOC/INSTR-*ima*, ACC-*e*, VOC-*i*, p. ex. *prozor*-NOM.SG.M 'fenêtre': *prozori*-NOM.PL.M, *prozora*-GEN.PL.M, *prozorima*-DAT.PL.M), tandis que les noms avec une seule syllabe ajoutent l'extension (suffixe) *-ov/-ev-* avant les désinences (p. ex. *stol*-NOM.SG.M 'table': *stolovi*-NOM.PL.M, *stolova*-GEN.PL.M, *stolovima*-DAT.PL.M). Cela fonctionne pour la grande majorité des noms, bien qu'il y ait des exceptions, comme le nom *pas*-NOM.SG.M ('chien') qui se réalise comme *psi*-NOM.PL.M au lieu de \**pasovi*, ou nom *tigar*-NOM.SG.M ('tigre') qui devient *tigrovi*-NOM.PL.M au lieu de \**tigri*. Ces exceptions peuvent être liées à certains traits phonologiques. Ce qui nous intéresse ici évidemment, ce sont les cas où un seul nom peut avoir deux formes au pluriel. Il existe plus de 80 lexèmes qui peuvent réaliser les deux types de pluriel — le pluriel dit *court* (sans *-ov/-ev-*) et le pluriel dit *long* (avec *-ov/-ev-*) — comme dans les exemples de noms dans Figure 5, respectivement *galeb*-NOM.SG.M 'mouette' > *galebi/galebovi*-NOM.PL.M ; *kesten*-NOM.SG.M 'maron' > *kesteni/kestenovi*-NOM.PL.M ; *korijen*-NOM.SG.M 'racine' > *korijeni/korijenovi*-NOM.PL.M ; *ležaj*'lit' > *ležaji/ležajevi*-NOM.PL.M ; *pauk*-NOM.SG.M 'araignée' > *pauci/paukovi*-NOM.PL.M ; *plamen*-NOM.SG.M 'flamme' > *plameni/plemenovi* ; *pojas*-NOM.SG.M 'ceinture' > *pojasi/pojasevi*-NOM.PL.M ; *prsten*-NOM.SG.M 'bague' > *prsteni/prstenovi*-NOM.PL.M ; *slučaj*-NOM.SG.M 'cas' > *slučaji/slučajevi*-NOM.PL.M. Les proportions des deux formes pluriels dans *forum.hr* (Ljubešić & Klubička 2014) pour un nombre limite des noms a été vérifiée, ce qui est présenté dans Figure 3 (v. Bošnjak Botica *et al.* 2022b).

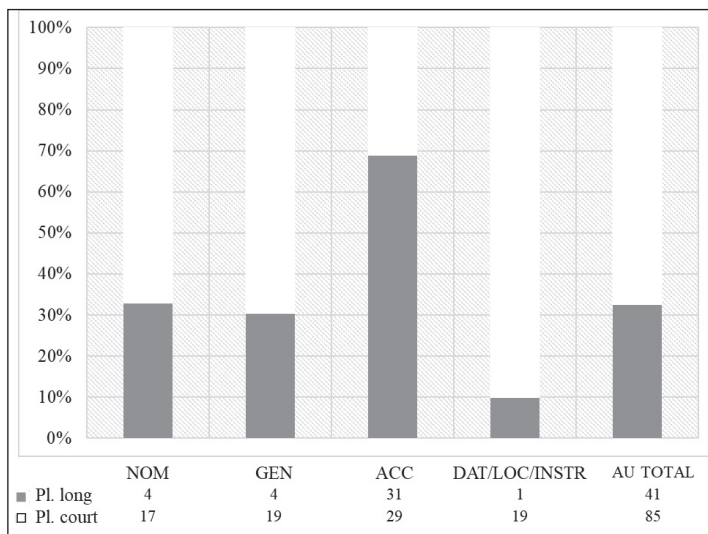
En examinant les différentes formes d'un nom, nous pouvons observer la distribution des deux formes dans différents cas du paradigme. Prenons par exemple le nom *palac*-NOM.SG.M 'pouce' dont les formes plurielles sont *palci/palčevi*-NOM.PL.M. Comme le montre la Figure 4, le pluriel court est dominant dans *Riznica* (Ćavar & Brozović Rončević 2012), sauf pour le cas accusatif, où l'utilisation de la forme longue est constatée (v. Bošnjak Botica *et al.* 2022b). Les attestations nous indiquent qu'il s'agit ici d'une locution idiomatique *držati*

*palčeve* ‘garder les doigts croisés’ où la forme longue est utilisée. Le nombre élevé de cette expression augmente significativement la distribution totale de la forme longue pour ce nom, ce qui met en question la nature du doubletisme dans des cas pareils. Notre prochaine recherche vise à découvrir le modèle qui s’impose lors de la formation du pluriel de ce groupe de noms, en essayant de prévoir les tendances futures.

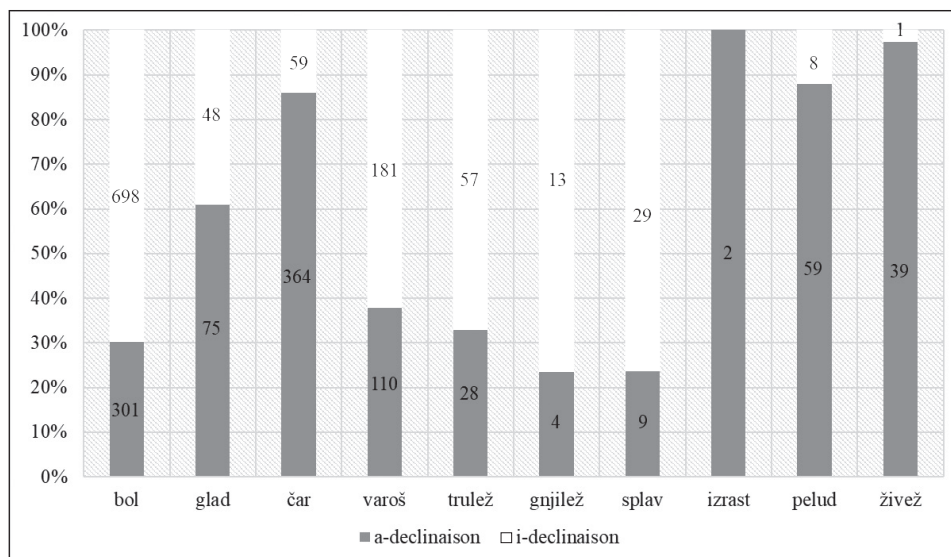
**Figure 3.** Les proportions des deux formes de noms pluriels masculins (le pluriel court vs. long) dans le corpus *HrWaC*



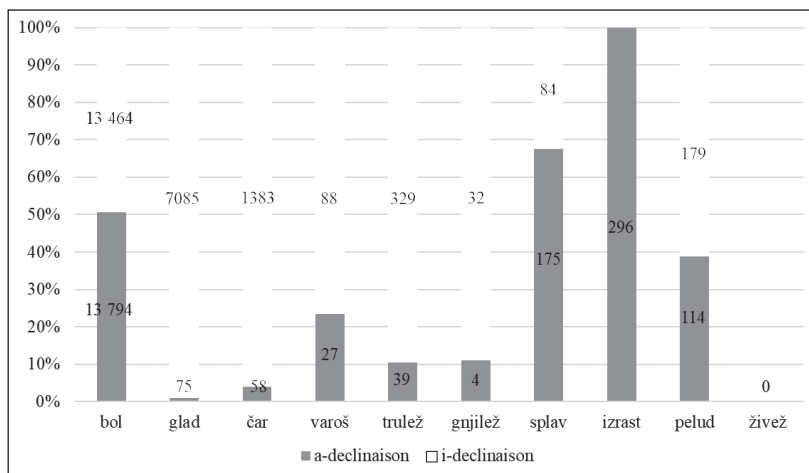
**Figure 4.** Les proportions des deux formes au pluriel masculin du nom *palac* ‘pouce’ (le pluriel court vs. long) dans le corpus *Riznica*



Les exemples suivants montrent comment les changements dans la distribution des déclinaisons (*a-déclinaison* et *i-déclinaison*) de certains noms dans les corpus des textes du genre littéraire (*Riznica*, Čavar & Brozović Rončević 2012) et non-littéraire (*hrWaC*, Ljubešić & Klubička 2014) peuvent être tracés au fil du temps. En comparant la distribution des noms dans *Riznica* (Figure 5) et *hrWaC* (Figure 6), nous pouvons remarquer que non seulement la distribution des deux genres a changé mais aussi certains noms, tels que le nom *glad* ‘famine’, n’apparaissent plus dans les deux genres et sont aujourd’hui réalisés uniquement au féminin (v. Bošnjak Botica *et al.* 2022b). Le nom *splav* ‘radeau’, par exemple, montre une augmentation de la forme masculine due, entre autres, à l’emploi dans le contexte des bars-restaurants flottants à Belgrade. Donc, en faisant la comparaison entre un corpus contenant les textes plus ou moins littéraires et un corpus contemporain, il est possible de tracer la direction du changement des unités linguistiques. Figure 5 montre la proportion des noms à deux déclinaisons (*a-déclinaison* et *i-déclinaison*) dans le corpus des textes littéraires (*Riznica*), alors que la Figure 6 le montre dans le corpus du genre non-littéraire (*hrWaC*).



**Figure 5.** Les proportions des noms à deux déclinaisons dans le corpus littéraire *Riznica*



**Figure 6.** Les proportions des noms à deux déclinaisons dans le corpus contemporain *hrWaC*

### 3. Conclusion

La recherche sur la surabondance des formes dans la flexion croate donne un aperçu de la richesse et de la complexité du système flexionnel de cette langue. Ce système est très proche aux systèmes d'autres langues slaves, notamment la langue serbe, la langue bosniaque et la langue monténégrine. Les formes concurrentes, c'est-à-dire les formes ayant le même sens grammatical mais des morphèmes flexionnels différents, sont présentes dans de nombreuses catégories de mots variables, telles que les noms, les pronoms, les verbes et les adjectifs. Ce phénomène remet en question le principe du blocage morphologique qui sous-entend une correspondance unique entre la forme et la fonction, autrement dit, qui postule que l'existence d'une forme empêche l'apparition et l'emploi d'autre forme ayant la même fonction. Au lieu de cela, la surabondance met en évidence l'existence de formes multiples qui peuvent coexister dans les mêmes contextes grammaticaux et sémantiques, sans différence fonctionnelle claire entre elles. L'étude des lexèmes surabondants dans la base de données des doublets morphologiques offre une contribution précieuse au débat sur la nature de la langue. Les données recueillies permettent de remettre en question l'idée selon laquelle la langue serait un système parfait caractérisé par une correspondance stricte entre forme et fonction.

Les formes doubles observées dans la langue croate révèlent leur présence régulière et non accidentelle. Il semble qu'elles ne soient pas simplement des



phénomènes éphémères ou des transitions entre différentes formes, mais plutôt des éléments intégrés et fonctionnels dans la structure linguistique du croate.

L'analyse diachronique et diatopique des formes doubles permettra de mieux comprendre leur origine, leur évolution et leur distribution géographique. Ces données suggèrent que la surabondance morphologique est normalement intralinguistique, résultant de l'évolution interne d'une langue, mais qu'elle pouvait être, dans certains cas qu'ils doivent être investigués plus profondément, influencée par des contacts linguistiques avec d'autres systèmes linguistiques proches.

En étudiant les formes concurrentes dans le contexte plus large de la morphologie et de la structure linguistique, nous sommes en mesure d'aborder des questions plus profondes sur la nature de la langue elle-même. Les formes doubles ou triples ne sont pas des anomalies, mais font partie intégrante d'un système linguistique morphologiquement plus riche, ce qui est évident pour les langues slaves. Ainsi, l'étude de ce phénomène offre des perspectives nouvelles pour comprendre la dynamique de la langue. Cette approche ouvre potentiellement la voie à une meilleure compréhension des variations linguistiques et des mécanismes qui sous-tendent la diversité des formes dans une langue donnée.

D'un point de vue pratique, les résultats de la recherche entreprise peuvent s'implémenter dans une description lexicographique plus précise.

## Références bibliographiques

- Aronoff 1976 : M. Aronoff, *Word Formation in Generative Grammar, Linguistic Inquiry*, Cambridge : The MIT Press.
- Bermel & Knittl 2012 : N. Bermel & L. Knittl, « Corpus frequency and acceptability judgements: A study of morphosyntactic variants in Czech », *Corpus Linguistics and Linguistic Theory*, n° 8(2), 241–275.
- Bermel *et al.* 2015 : N. Bermel, L. Knittl & J. Russell, « Morphological variation and sensitivity to frequency of forms among native speakers of Czech », *Russian Linguistics*, n° 39(3), 1–26.
- Bošnjak Botica & Hržica 2016 : T. Bošnjak Botica & G. Hržica, « Overabundance in Croatian dual-class verbs », *Fluminensia*, n° 28(1), 83–106.
- Bošnjak Botica *et al.* 2018 : T. Bošnjak Botica, J. Budja & G. Hržica, « Rodovna i sklonidbena dvojnost imenica u hrvatskom jeziku », *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, n° 44(1), 57–88.
- Bošnjak Botica *et al.* 2019 : T. Bošnjak Botica, D. Lečić, G. Hržica & J. Budja, « Baza hrvatskih morfoloških dubleta (DvojBa) », *Suvremena lingvistika*, n° 45(87), 95–108.
- Bošnjak Botica *et al.* 2022a : T. Bošnjak Botica, J. Polančec, G. Hržica, S. Košutar & M. Musulin, « Udio različitih nastavaka u genitivu množine imenica e-sklonidbe », in *Standardni i nestandardni idiomi: 36. međunarodni znanstveni skup Hrvatskog društva za primijenjenu lingvistiku*, Osijek, 9–11 juin 2022 (communication).



- Bošnjak Botica *et al.* 2022b : T. Bošnjak Botica, J. Polančec & R. Sviben, « Korpusno istraživanje hrvatskih imenica s dugom i kratkom množinom », *Jezikoslovlje*, n° 23(1), 35–74.
- Cappellaro 2013 : C. Cappellaro « Overabundance in diachrony: a case study », in S. Cruschina, M. Maiden & J. C. Smith (dirs), *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*, Oxford : Oxford University Press, 209–220.
- Carstairs-McCarthy 2010 : A. Carstairs-McCarthy, *The evolution of morphology*, Oxford : Oxford University Press.
- Ćavar & Brozović Rončević 2012 : D. Ćavar & D. Brozović Rončević, « Riznica: The Croatian Language Corpus », *Prace filologiczne*, n° 63, 51–65.
- Clark 1987 : E. V. Clark, « The principle of contrast: A constraint on language acquisition », in B. MacWhinney (dir.), *Mechanisms of language acquisition*, Hillsdale, NJ : Lawrence Erlbaum Associates, 1–33.
- Corbett 2007 : G. G. Corbett, « Canonical typology, suppletion and possible words », *Language*, n° 83(1), 8–41.
- Corbett 2009 : G. G. Corbett, « Canonical Inflectional Classes », in F. Montermini, G. Boyé & J. Tseng (dirs), *Selected Proceedings of the 6th Décembrettes*, Somerville, MA : Cascadilla Proceedings Project, 1–11.
- Cummins 1995 : G. M. Cummins, « Locative in Czech: “-U” or “-e”? Choosing Locative Singular Endings in Czech Nouns », *Slavic and East European Journal*, n° 39(2), 241–260.
- Erjavec *et al.* 2003 : T. Erjavec, C. Krstev, V. Petkevič, K. Simov, M. Tadić & D. Vitas, « The MULTTEXT-East Morphosyntactic Specifications for Slavic Languages », in T. Erjavec & D. Vitas (dirs), *Proceedings of the EACL2003 Workshop on Morphological Processing of Slavic Languages*, Budapest : Association for Computational Linguistics, 25–32.
- Fehringer 2004 : C. Fehringer, « How Stable are Morphological Doublets? A Case Study of ə and Ø Variants in Dutch and German », *Journal of Germanic Linguistics*, n° 16(4), 285–329.
- Jelaska & Bošnjak Botica 2019 : Z. Jelaska, & T. Bošnjak Botica, « Conjugational Types in Croatian », *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, n° 45(1), 47–74.
- Lečić 2015 : D. Lečić, « Morphological doublets in Croatian », *Russian Linguistics*, n° 39(3), 375–393.
- Ljubešić & Klubička 2014 : N. Ljubešić & F. Klubička, « {bs,hr,sr}WaC — Web Corpora of Bosnian, Croatian and Serbian », in F. Bildhaueret & R. Schäfer (dirs), *Proceedings of the 9th Web as Corpus Workshop (WaC-9) @ EACL 2014*, Göteborg : Association for Computational Linguistics, 29–35.
- Mörth & Dressler 2014 : K. Mörth & W. U. Dressler, « German plural doublets with and without meaning differentiation », in F. Rainer, F. Gardani, H. C. Luschützkyet & W. U. Dressler (dirs), *Morphology and Meaning: Selected papers from the 15th International Morphology Meeting, Vienna, February 2012*, Amsterdam : John Benjamins, 249–258.

- Naghzgy-Kohan & Kuteva 2016 : M. Naghzy-Kohan & T. Kuteva, « On competition and blocking in inflectional morphology: Evidence from the domain of number in New Persian », *Folia Linguistica*, n° 50(1), 65–96.
- Santilli 2014 : E. Santilli, *Italian Comparatives: a Case of Overabundance?* [Thèse de doctorat non publiée], Université de l'Aquila.
- Thornton 2011 : A. M. Thornton, « Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology », in M. Maiden, J. C. Smith, M. Goldbachet & M.-O. Hinzelin (dirs.), *Morphological autonomy: perspectives from Romance inflectional morphology*, Oxford : Oxford University Press, 358–381.
- Thornton 2012 : A. M. Thornton, « Reduction and maintenance of overabundance: A case study on Italian verb paradigms », *Word Structure*, n° 5(2), 183–207.

Tomislava Bošnjak Botica i Sara Košutar

### Preobilje izraza u hrvatskoj fleksiji

Po načelu jezične ekonomičnosti, u jeziku se nastoji izbjeći gomilanje izraza istoga značenja, što u načelu funkcionira tako da se jednomu izrazu pridružuje jedno značenje odnosno funkcija. Međutim, služeći se jezikom svakodnevno, znamo da se to načelo ne provodi dosljedno i da jednomu značenju nerijetko odgovara više izraza. U ovom se radu pozornost usmjerava upravo na pojavu dvaju ili više izraza za istu gramatičku funkciju unutar hrvatske fleksijske morfologije. Navedena se pojava u lingvistici naziva *preobilje* (engl. *overabundance*, fr. *surabondance*) i u posljednje vrijeme izaziva veliko zanimanje istraživača, pogotovo u jezicima koji su manje istraživani, poput hrvatskoga i drugih slavenskih jezika jer se pokazalo da *učinak blokiranja* (engl. *blocking effect*, fr. *blocage*) ipak negdje ne djeluje. Iznose se primjeri preobilja u hrvatskom jeziku (npr. prezent glagola *skitati se*: *skitam/skićem se*, instrumental imenice *strast*: *strasti/strašću*, genitiv množine imenice *kruška*: *krušaka/kruška/kruški*, cijela množina imenice *palac*: *palci/palčevi*, komparativ pridjeva *dug*: *duži/dulji*, genitiv zamjenice *njegov*: *njegova/njegovog* itd.). Gledajući ukupan broj flektivnih riječi u hrvatskom jeziku, posrijedi je rubna pojava, ali je ona tipološki važna te hrvatski, zajedno s nekim slavenskim i ugrofinskim jezicima, smješta u skupinu jezika s visokim stupnjem morfološke dubletnosti. Predstavlja se i *Baza hrvatskih morfoloških dubleta* kao jezični resurs dostupan ne samo znanstvenoj zajednici za dublje proučavanje ove pojave nego i tzv. običnim korisnicima jezika koji će njome moći provjeriti status pojedinoga oblika. Na primjeru triju kategorija pokazali smo kako izgleda distribucija pojedinih oblika u dostupnim korpusima te kako se udio pojedinih oblika može s vremenom mijenjati.

*Gljučne riječi*: morfološko preobilje, baza dubleta, fleksija, hrvatski jezik.