PROJECT: Towards a typology of clausal prolepsis (TTCP)

STATE OF THE ART, PROBLEM STATEMENT, AND OBJECTIVES
There are two main ways of integrating two clauses into one utterance: they can be coordinated, as in (1a), or one clause can be integrated into the other, as in (1b), where the clause *that Bill sleeps* serves as the direct object of the verb *thinks* in the main clause.

(1)   a.   John eats and Bill sleeps.
      b.   John thinks that Bill sleeps.

In the languages of the world, the pattern in (1b) is rarer than the one in (1a). Many languages have developed alternative strategies to avoid clauses from being embedded into other clauses. One such strategy are so-called **proleptic proforms**. They form the focus of the TTCP-project.

In the examples in (2), the clause *that John will win* is heralded by the presence of *it*. This pronoun occupies the position reserved for the clause—the subject of *is likely* in (2a) and the object of the preposition *of* in (2b)—while the clause that we would expect to be embedded (*that John will win*) occupies a (right-)peripheral position.

(2)   a.   **It** is likely that John will win.
      b.   I am sure of **it** that John will win.

Given the important role proleptic proforms play in a process as central and fundamental to natural language as clausal embedding, they have received ample attention in the formal linguistic literature of the last six decades. Despite the sizeable body of work, however, there is little agreement as to what is the correct analysis of this phenomenon. The existing analyses can be split up into three main subtypes: (a) a first approach takes pronouns like *it* in (2) to be **meaningless**, i.e. semantically vacuous placeholders for their accompanying clauses (see e.g. Postal and Pullum (1988), Williams (1994), Chomsky (1995)), (b) a second one argues them to be **arguments**; from this perspective, *it* in (2a) is the real subject of the sentence and the real object of the preposition in (2b), with the clause *that John will win* occupying a modifier-like position (e.g. Bennis (1986), Ruys (2010)), (c) the third option is to treat them as **predicates**; in this case, the subject of *is likely* in (2a) and the object of the preposition *of* in (2b) would be something like *that John will win is it* (e.g. Moro (1997)).

While the literature on proleptic proforms has developed a rich body of tests and criteria to distinguish between the three analyses, there is little consensus as to which analysis is correct, and as a result, the research has stagnated in recent years. One of the main reasons behind this lack of progress, we believe, is the **overly Anglo-centric perspective** of much of the literature. This narrowing of the empirical domain has left one aspect of proleptic proforms systematically underexplored: their morphology. While in English proleptic proforms are typically personal pronouns like *it*, other languages have a richer inventory. Examining how those other types of proforms behave with respect to the tests developed in the literature is likely to shed new light on proleptic proforms and could provide a new impetus for the research into the phenomenon of prolepsis and into clausal embedding more generally. **The TTCP-project** takes precisely this tack: it will be **the first to systematically examine the interaction between the morphology of proleptic proforms and their syntactic characteristics**.

In order to achieve this goal, the TTCP-project focuses on two languages that show morphological variation in their proleptic proforms: **Dutch and Hungarian**. As the examples in (3) show, Dutch can use either the personal pronoun *het* 'it' or the locative proform *er* 'there' in proleptic position (Bennis 1986, Ruys 2010), while Hungarian shows variation between the personal/demonstrative pronoun *az* 'it/that' and the manner adverb *úgy* 'so/thus', as shown in (4) (den Dikken 2018).

(3)   { **Het** / **Er**   } is gebleken   dat 14% van de Belgen   in armoede leeft.
        it   there   is turned.out that 14% of   the Belgians in poverty   lives
        'It has turned out that 14% of the Belgian population lives in poverty.'   Dutch

(4)   { **Az** / **Úgy** } látszik, hogy Mari terhes.
        it   so   seems that Mari pregnant
        'It seems that Mari is pregnant.'   Hungarian

This means the combined empirical domain of Dutch and Hungarian provides an ideal testing ground for

examining the interaction between the morphology of clausal prolepsis and its syntactic properties, thus furthering our insight into this phenomenon and into clausal embedding more generally. How exactly this will proceed is outlined in the next section.

METHODOLOGY

As pointed out above, the existing literature on proleptic proforms has yielded a large battery of tests and criteria for determining the syntactic status of a proleptic proform and hence the corresponding syntactic analysis (see a.o. Postal and Pullum (1988), Bennis (1986), Ruys (2010), Moro (1997)). This list of tests includes the following **ten syntactic configurations**: (a) the ability of proleptic proforms to bind reflexive pronouns, (b) the possibility of coordinating proleptic proforms, (c) the possibility of proleptic proforms to occur in nominalizations, (d) the ability of proleptic proforms to act as subject in a so-called *tough*-construction, (e) the ability of proleptic proforms to bind the implicit subject of an infinitival clause, (f) the ability of proleptic proforms to bind parasitic gaps, (g) the possibility of extraction out of the clause associated with the proleptic proform, (h) the obligatoriness of the proleptic proform, (i) the possibility of adding a copula to the embedded clause, and (j) the ability of proleptic proforms to be questioned.

When combined, these ten criteria maximally distinguish between the three analyses of proleptic proforms. This is shown in the table in (5). For example, with respect to context (a) (the ability of proleptic proforms to bind reflexive pronouns), the analysis of proleptic proforms as meaningless elements predicts this to be ill-formed, as does the predicate analysis, while the argument analysis predicts such cases to be grammatical. As the table shows, the combination of the ten configurations creates a unique empirical profile for each analysis, thus serving as a litmus test for distinguishing between the three analyses.

(5)	analysis of proleptic proforms	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
	1: meaningless elements	*	*	*	*	*	*	✓	*	*	*
	2: arguments	✓	✓	✓	✓	✓	✓	*	✓	*	✓
	3: predicates	*	✓	*	*	*	*	*	✓	✓	✓

An important novelty of the TTCP-project concerns the fact that it will map the criteria in (5) against the morphology of proleptic proforms: **the project will systematically test these ten criteria for each of the four configurations in (3)–(4)**. A unique feature of the data in (3)–(4) is the fact that the triple *there/it/so* represents the prototypical instantiations of the three analyses outlined above: *there* is the meaningless element we also find in impersonal passives like (6), *it*, as a nominal, is a prototypical argument, and *so* is a prototypical predicate (Corver and Thiersch 2001).

(6) Er is gelachen.
    there is laughed
    'People laughed.'   Dutch

This means that the TTCP-project will be testing the prototypical contexts for detecting the three analyses of proleptic proforms against the prototypical morphological instantiations of those analyses. **The main hypothesis of the TTCP-project is that the three analyses of clausal prolepsis stand side-by-side, rather than in complementary distribution**: it is not a matter of which analysis is correct in absolute terms, but rather which one is correct in which context. The insights gleaned from Dutch and Hungarian can then be transferred to morphologically poorer languages like English, where the form of the proleptic proform offers no cues as to its underlying analysis.

The TTCP-project will test this hypothesis through **large-scale questionnaires distributed among native speakers of Dutch and Hungarian**. The Dutch questionnaires will be distributed using Taalpost (https://onzetaal.nl/nieuws-en-dossiers/nieuwsbrieven/taalpost/), which is the newsletter of the Dutch Language Society, the blog neerlandistiek.nl (https://www.neerlandistiek.nl), and the informant network of the Meertens Institute in Amsterdam, which comprises 7,000 informants. For Hungarian, there are no such informant networks. Instead, we will make use of the author team of *The Syntax of Hungarian*, a multi-authored grammar project, the first two volumes of which have appeared in the series *Comprehensive Grammar Resources* with Amsterdam University Press. The team leader of this project is prof. István Kenesei (Hungarian Academy of Sciences) and the rest of the team consists of an elaborate network of

linguists across the whole of Hungary, who are experienced in filling out questionnaires.

The results of these questionnaires will be processed using basic descriptive statistical methods and will then serve as input for formal theoretical analyses of proleptic proforms.

ANTICIPATED RESULTS AND DELIVERABLES

The TTCP-project intends to achieve the following goals: (1) a detailed empirical overview of the distribution of proleptic proforms across ten different syntactic contexts and across the Dutch- and Hungarian-speaking area, (2) a formal theoretical analysis of the distinction between *het* 'it' and *er* 'there' in Dutch proleptic constructions, (3) a formal theoretical analysis of the distinction between *az* 'it/that' and *úgy* 'so/thus' in Hungarian proleptic constructions, (4) the foundations of a general theory of proleptic proforms that is able to predict which proleptic strategy (meaningless/argument/predicate) is used in which context and as a result, a deeper understanding of and insight into clausal embedding, and (5) a stepping stone towards a larger, more ambitious project application focusing on prolepsis and argumenthood in the languages of Europe.

Each of these anticipated results comes with its own deliverable. The empirical-methodological side of the TTCP-project will yield a large Dutch-Hungarian **database of native speaker judgments**. This database will be unique in that it is the first to systematically test a large number of syntactic characteristics of proleptic proforms with a wide range of native speakers, and moreover, in doing so in a way that allows for direct comparison between two typologically unrelated languages, i.e. Dutch and Hungarian. This database, which will be of great value to other researchers interested in these or related topics, will be made available online for the broader research community at the end of the TTCP-project.

The formal theoretical analyses of Dutch and Hungarian proleptic constructions will lead to **two journal publications** in an A-list journal like *Natural Language and Linguistic Theory*, *Linguistic Inquiry*, *Glossa*, or *Syntax*, one in which the Belgian half of the project team takes the lead and one in which the Hungarian half does. The general theory of proleptic proforms and clausal embedding will lead to an **overview article** by the entire project team in *Language and Linguistics Compass*, and in order to prepare for the follow-up project—for more details, see attachment 4—the TTCP-project will organize an **international workshop** on prolepsis and argumenthood at the end of the second year of the project. This workshop will be used to (a) present the main results of the TTCP-project to an international audience of experts, (b) extend and strengthen the international network of the project team by inviting strategic partners as invited speaker, and (c) create international awareness and attention for the topic of prolepsis and argumenthood through an open call for papers.

PROJECT TEAM, WORK PACKAGES, AND TIMING

The Belgian half of the project will consist of Jeroen van Craenenbroeck (PI) and a part-time (50%) postdoc, who will be recruited at KU Leuven for two years. These two researchers will form a part of the Center for Research in Syntax, Semantics, and Phonology (CRISSP, http://www.crissp.be) at KU Leuven and thus benefit from the expertise available within that center: Guido Vanden Wyngaerd (morphology), Dany Jaspers (formal semantics and dialectology), Hans Smessaert (formal semantics and Dutch grammar), and Cora Pots (morphosyntactic variation).

The Hungarian half of the project will consist of Marcel den Dikken (PI) and Krisztina Szécsényi (PI). In addition, at ELTE the project will benefit from the presence of Mark Newson, a senior syntactician whose interest in constraint-based theoretical frameworks (Optimality Theory and his own developments thereof) will provide a welcome different perspective on the questions addressed in this project. Furthermore, the participation of researchers at the Research Institute for Linguistics of the Hungarian Academy of Sciences (to which Den Dikken is also affiliated), specifically Hans-Martin Gaertner (syntax, pragmatics), Katalin É. Kiss (syntax, semantics, morphology), Balázs Surányi (syntax), and Huba Bartos (syntax), will provide an essential platform for this project.

The workload of the project is divided into the following four work packages:

WORK PACKAGE #1: literature review and questionnaires (postdoc + PI Szécsényi, in collaboration with the other two PIs, month 1–8)
- review of the linguistic literature on proleptic proforms
- construction of the questionnaires
- distribution of the questionnaires among native speakers and collection of the results

- construction of the database
- application of basic descriptive statistical tests to the results in the database

WORK PACKAGE #2: formal theoretical analysis of Dutch and Hungarian proleptic constructions (entire team, month 9–15)
- theoretical analysis of the empirical results of the questionnaires
- presentation of intermediate stages of the analysis at national (BKL, De Grote Taaldag, BLINC) and international (CGSW, GLOW, NELS, ICSH) conferences
- preparation of a journal publication

WORK PACKAGE #3: towards a unified theory of prolepsis and clausal embedding (entire team, month 16–22)
- integration of the theoretical analyses of Dutch and Hungarian proleptic proforms
- development of a unified theoretical analysis of prolepsis and clausal embedding
- presentation of intermediate stages of the analysis at national (BKL, De Grote Taaldag, BLINC) and international (CGSW, GLOW, NELS, ICSH) conferences
- preparation of an overview article

WORK PACKAGE #4: project management and workshop (the three PIs, month 1-24)
- two project meetings: one in Leuven at the beginning of the questionnaire stage (December 2019), to discuss the construction of the questionnaires, and one in Budapest (April 2021) to discuss and develop the unified theory of prolepsis and clausal embedding
- workshop organization (April–September 2021): presentation of the results of the TTCP-project, invitation of external experts, distribution of an open call for papers and abstract evaluation
- regular Skype-sessions between the postdoc and PI Szécsényi (in the context of WP1) or the entire project team (in the context of WP2 and WP3) to discuss progress and potential challenges, and to guarantee intra-project compatibility and coherence
- preparation of the follow-up application for European funding (April–September 2021): contacts with European partners, development of the proposal
- budget management

	2019			2020												2021								
	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9
WP#1a: literature review	▓	▓																						
WP#1b: questionnaires			▓	▓	▓	▓	▓	▓																
WP#2: analyses of Dutch & Hungarian									▓	▓	▓	▓	▓	▓	▓									
WP#3: unified theory of prolepsis																▓	▓	▓	▓	▓	▓	▓		
WP#4a: project management	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓
WP#4b: preparation follow-up application																			▓	▓	▓	▓	▓	▓
Milestones and deliverables																								
- database								⊙																
- article publications															⊙⊙							⊙		
- conference presentations												⊙		⊙					⊙		⊙			
- workshop																								⊙
- project meetings			⊙																⊙					

References

Bennis, Hans. 1986. *Gaps and dummies*. Dordrecht: Foris Publications.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press.

Corver, Norbert, and Craig Thiersch. 2001. Remarks on parentheticals. In *Progress in grammar: Articles at the twentieth anniversary of the Comparison of Grammatical Models Group in Tilburg*, ed. Marc van Oostendorp and Elena Anagnostopoulou. Amsterdam: Meertens Institute.

den Dikken, Marcel. 2018. *Dependency and directionality*. Cambridge: Cambridge University Press.

Moro, Andrea. 1997. *The raising of predicates : predicative noun phrases and the theory of clause structure*. Cambridge: Cambridge University Press.

Postal, Paul, and Geoffrey Pullum. 1988. Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19:635–679.

Ruys, E. G. 2010. Expletive selection and CP arguments in Dutch. *The Journal of Comparative Germanic Linguistics* 13:141–178.

Williams, Edwin. 1994. *Thematic structure in syntax*. Cambridge, Massachusetts: MIT Press.