Headless relative clauses and the syntax-semantics mapping:
Evidence from Mesoamerica

IVANO CAPONIGRO
University of California San Diego

1. Introduction

A familiar component of the syntax-semantics mapping goes as follows. Tensed full clauses are morpho-syntactic devices that either convey propositional content, as it is the case for many matrix clauses and embedded clauses, or act as nominal modifiers, as it is the case for headed relative clauses. On the other hand, full nominal expressions, i.e., DPs, are the constituents that are used to refer to individuals (including kinds) or quantify over sets of individuals by denoting (a subset of) generalized quantifiers (i.e., functions from sets of individuals to truth-values).¹

Although much less studied, there’s a broad family of embedded full tensed clauses that exhibit a semantic behavior resembling those of DPs: they don’t convey propositional content, but rather denote individuals or generalized quantifiers. For instance, the wh- clause what Pablo made in the sentence Frida ate what Pablo made is referring to the food that Pablo made, like the referential DP the food by Pablo, even if the clause occurs without a definite article (a D head) or a noun (an N head). Some of these clauses can occur as complements of determiners, which is a position that is usually reserved for NPs. For instance, the string most that like Pablo with the underlined full tensed clause following the D most denotes the same type of generalized quantifier as the quantificational DP most people. We label this family of clauses headless relative clauses (henceforth, [−H]RCs).

In this paper, we show that the study of the semantic behavior of [−H]RCs leads to the generalization about the syntax-semantics mapping for natural languages that is given in (1).

(1) Generalization about the syntax-semantics mapping in natural languages

Natural languages can systematically use clauses in the form of [−H]RCs to denote individuals (including kinds) without any specific morpho-syntactic marking. On the other hand, quantification via any type of [−H]RC requires an overt specific morpho-syntactic marker; different kinds of markers can be employed within and across languages.

For instance, the [−H]RC what Pablo made is a plain wh- clause without any extra morpho-syntactic marking and can only be interpreted as referring to the things that Pablo made—like a definite DP. It cannot be equivalent to a quantificational DP like some/most/all/none of the things that Pablo made, regardless of the context. On the other hand, a [−H]RC like most that like Pablo is interpreted as a generalized quantifier of the same type as the one that is denoted by the

¹ Although Montague (1973) shows that both definite and quantificational DPs can be analyzed as denoting generalized quantifiers, Partee (1987) provides arguments to allow definite DPs (unlike quantificational DPs) to denote individuals as well.
quantificational DP *most people*. The presence of an overt morpho-syntactic maker—the D head *most*, in this case—determines the kind of meaning that is assigned to the whole [−H]RC.

We support this new generalization with new findings from a collaborative project that has systematically studied [−H]RCs in fifteen languages from five language families, all from Mesoamerica but one. We also provide a compositional semantic analysis for each of the five main kinds of [−H]RCs we have identified. For some, our analysis is the first ever. For the others, analyses already exist for other languages; we show that our findings are compatible with those analyses and bring further support to them.

To the best of our knowledge, the generalization in (1) holds for and is supported by not only our findings but also the findings about [−H]RCs in all the languages in which they have been investigated so far. Since most of the previous studies have focused on just one kind of [−H]RC in one language, it was difficult for them to detect and highlight this general pattern. Therefore, we have chosen to investigate all the main varieties of [−H]RCs in a systematic, consistent, and comparable way across languages and language families.

The remainder of the paper is structured as follows. Section 2 briefly introduces Mesoamerican languages, our project, and the languages we have investigated. Section 3 provides a general definition of [−H]RC. Each of the Sections 4–6 introduces a different kind of [−H]RC, summarizes the main findings from our crosslinguistic investigation, and provides a semantic analysis for each kind of [−H]RC. In particular, Section 4 deals with free relative clauses and its three main sub-varieties (Sections 4.1–4.3), Section 5 discusses light-headed relative clauses, and Section 6 presents super-free relative clauses. Section 7 concludes with some broader considerations.

2. Mesoamerican languages and our project

The term *Mesoamerica* refers to both a geographical region and a cultural and linguistic area. The region extends “from the Pánuco River in northern Mexico to the Lempa River in El Salvador, but also includes the Pacific coast of Nicaragua and Costa Rica” (Campbell 1997: 156). The indigenous populations from this region share a large number of cultural and linguistic traits (Campbell et al. 1986). Mesoamerican languages form a “group of more than 125 languages classified into some 10 language families (including language isolates) that are native to Mesoamerica,” including Uto-Aztecan, Oto-Manguean, Mayan, and Mixe-Zoquean. (Campbell 2014). Word order restrictions are among their common features: no Mesoamerican language exhibits verb-final basic word order, “although Mesoamerica is bounded by languages to both the north and the south that have SOV basic word order” (Campbell 2014).

Previous work on Mesoamerican languages had shown evidence of [−H]RCs (Caponigro 2020 and reference therein). Our project has investigated [−H]RCs systematically and in depth in fifteen languages, all from Mesoamerica but one. The languages are listed in (2) together with the names of the scholars who investigated them. We refer to Caponigro et al. (2020) for all relevant data and the language-specific descriptions.
Languages investigated in Caponigro et al (2020) and the scholars who investigated them:

**Mesoamerican**

UTO-AZTECAN
- *Southeastern Tepehuan (O’dam)*, by Gabriela García Salido
- *Tlaxcala Nahuatl*, by Lucero Flores-Nájera

OTO-MANGUEAN
- *Acazulco Otomi*, by Néstor Hernández-Green
- *Matlatzinca*, by Enrique Palancar and Leonardo Carranza Martínez
- *Iliatenco Me’phaa*, by Philip T. Duncan and Harold Torrence
- *San Pedro Mixtepec Zapotec*, by Pafnuncio Antonio-Ramos

MAYAN
- *K’iche’*, by Telma Angelina Can Pixabaj
- *Q’anjob’al*, by Eladio Mateo Toledo
- *Ch’ol*, by Justin Royer
- *Tsotsil and Tzeltal (Tzeltalan)*, by Gilles Polian and Judith Aissen
- *Yucatec Maya*, by Scott AnderBois and Miguel Oscar Chan Dzul

MIXE-ZOQUEAN
- *Sierra Popoluca*, by Wendy López Márquez

**Non-Mesoamerican**

CHIBCHAN
- *Pesh*, by Claudine Chamoreau

Our project has involved twenty-one scholars from Mexico, USA, Canada, and France and two preparatory workshops. The first workshop allowed all the scholars to share and agree on the same definitions of [−H]RCs, the methodology for the data collection, and the template to follow in organizing and presenting the data. The second workshop focused on presenting the data from the fieldwork and providing collective and individual feedback in preparation for the writing of the individual chapters. More details on its structure and organization can be found in Caponigro et al (2020: Preface).

3. **Definition of headless relative clauses**

We define [−H]RCs as a family of constructions sharing the properties in (3).

(3) **Properties characterizing [−H]RCs:**

I. They are embedded clauses;

II. They have an argument or an adjunct missing (it can sometimes be marked by a resumptive pronominal form);
III. They lack an “external nominal head”—a nominal head that precedes or follows them and is linked\(^2\) to the missing constituent;

IV. They exhibit the same distribution and interpretation as DPs or PPs.

Examples of \([-\text{H}]\)RCs are given in brackets in (4)–(6).

(4) Frida was very friendly with \([[-\text{H}]\text{RC} \text{those} \text{ Susana invited } __ \text{ from Mexico}]\).

(5) Paloma buys \([[-\text{H}]\text{RC} \text{whatever painting} \text{ Frida makes } __ ]\).

(6) Mercedes lives \([[-\text{H}]\text{RC} \text{where} \text{ she was born } __ ]\).

The \([-\text{H}]\)RCs in (4)–(6) share the four properties in (3). They are all embedded clauses inside a matrix clause (Property I). They all have a missing constituent, marked by the underscore (Property II): in (4) and (5), what is missing is the object of the transitive predicates invited and makes, while in (6) it is the locative adjunct to the intransitive predicate was born. None of the \([-\text{H}]\)RCs is preceded or followed by a nominal head (Property III): the \([-\text{H}]\)RC in (4) is introduced by the italicized demonstrative those, which crucially occurs without a nominal complement; the \([-\text{H}]\)RC in (5) is introduced by the italicized complex wh-phrase whatever painting, whose noun painting is part of the wh-phrase and inside the \([-\text{H}]\)RC, rather acting as an external head;\(^3\) the \([-\text{H}]\)RC in (6) is introduced by the italicized wh-word, which sits in the left peripheral position of the \([-\text{H}]\)RC. Last, the \([-\text{H}]\)RCs in (4), (5), and (6) can all be replaced and paraphrased by a DP or a PP (Property IV): the \([-\text{H}]\)RCs in (4) and (5) have the same distribution and interpretation as the bracketed DPs in (7) and (8), respectively; the bracketed string in (6), by contrast, shows an example of a \([-\text{H}]\)RC with the same distribution and interpretation as the bracketed PP in (9).

(7) Frida was very friendly with \([\text{DP} \text{the people that Susana invited} \text{ from Mexico}]\).

(8) Paloma buys \([\text{DP} \text{any} \text{ painting that Frida makes}]\).

(9) Mercedes lives \([\text{PP} \text{in the place where she was born}]\).

We summarize and further specify the properties in (3) with the syntactic schema in (10)a and the feature bundle in (10)b, which we use as a concise devise to define the different varieties of \([-\text{H}]\)RCs in the next sections as well.

(10) **Summary of the properties characterizing \([-\text{H}]\)RCs:**

a. \([(\text{DET}) [(\text{CP}) (\text{wh-/REL/COMP}) \text{ …} \text{...}]]\text{DP/PP}^4\]

b. \([\pm \text{D}, \pm \text{N}, \pm \text{WH}]\)

\(^2\)“Linked” is used here as a general label for any kind of morpho-syntactic (e.g., gender or number agreement) or semantic connection.

\(^3\)See Caponigro (2019) for data and arguments from English and other languages supporting this claim.

\(^4\)The subscript \text{DP/PP} occurring at the far right of the syntactic schema in (10)a does not indicate the actual syntactic categories of \([-\text{H}]\)RCs, but indicates the distributional similarities of \([-\text{H}]\)RCs with actual DPs and PPs.
[-H]RCs have no nominal head (−N). Some have a “light head,” which we have assigned the label Determiner (±D) for convenience and whose nature we shall discuss later. [-H]RCs can feature a wh-expression\(^5\) (wh-) from the set of those introducing wh- interrogative clauses with possible extra morphological marking (±WH), a non-wh relativizer (REL) of the same kind as those introducing headed relative clauses (including non-wh relative pronouns), a general complementizer (COMP) of the same kind as those introducing complement clauses (and maybe headed relative clauses too), a combination of those, or no marking at all. All [-H]RCs have a missing constituent ( _ ).

Despite the similarities highlighted in (3) and (10), [-H]RCs exhibit differences in interpretation, distribution, and/or morpho-syntactic properties that are discussed in the next sections. Overall, the label “headless relative clauses ([−H]RCs)” should be taken as a way to identify not just one single construction, but a cluster of related constructions, and to distinguish them from another cluster of related but different constructions—headed relative clauses. In particular, we have identified three main varieties of [-H]RCs: free relative clauses, light-headed relative clauses, and super-free relative clauses, based on the properties and feature bundle in (10), as shown in the schema in (11). Each kind, its label, and its features are discussed in the next sections.

(11) Varieties of [-H]RCs and their characterizing features:

\[
\begin{align*}
\text{[-H]RC} & \quad \text{Free Relative Clause (FR)} \\
& \quad [\text{-D, N, ±WH}] \\
& \quad [\text{CP wh-(COMP) _}]_{DP/PP}
\end{align*}
\]

\[
\begin{align*}
\text{[-H]RC} & \quad \text{Light-Headed Relative Clause (LHR)} \\
& \quad [\text{+D, N, ±WH}] \\
& \quad [\text{Det [CP (wh-/REL/COMP) _]}]_{DP/PP}
\end{align*}
\]

\[
\begin{align*}
\text{[-H]RC} & \quad \text{Super-Free Relative Clause (SFR)} \\
& \quad [\text{-D, N, -WH}] \\
& \quad [\text{CP (REL/COMP) _}]_{DP}
\end{align*}
\]

[-H]RCs are extremely productive in the languages we have investigated, as shown in Table 1.

\(^5\) We use the term wh-expression to refer both to a single wh-word occurring on its own (e.g., who, where, when, etc.) as well as to a wh-phrase that consists of a wh-word and other material, like a complement (e.g., what book, how much food) or a preposition (e.g., by means of what, with which friend) or both (e.g., together with how many other people).
Table 1. Productivity of varieties of \([-H]\)RCs in the languages under investigation\(^6\)

<table>
<thead>
<tr>
<th></th>
<th>UTO-AZT</th>
<th>OTO-MANGUEAN</th>
<th>MAYAN</th>
<th>MI-ZO</th>
<th>CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Te</td>
<td>Ú</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Na</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AO</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ma</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>IM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SZ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Kꞌ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Qꞌ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CJ</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CL</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TT</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>YM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pk</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LHR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SFR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: Marks: ✓: attested; *: not attested; ?: unclear if attested.

In the next sections, we discuss each kind of \([-H]\)RC in turn, highlighting the main criteria that we used to identify them. We also introduce further distinctions within each kind. Free relative clauses are discussed next (Section 4), then light-headed relative clauses (Section 5), and finally super-free relative clauses (Section 6).

4. Free relative clauses

We use the label free relative clauses (henceforth, FRs) for those \([-H]\)RCs that have no D head and are introduced by (a subset of) the wh-expressions in wh-interrogative clauses, with or without extra morpho-syntactic marking. In some of our languages, the wh-expression may co-occur with a complementizer. The characterizing features FRs are summarized by the syntactic schema in (12)a and the feature bundle in (12)b.

\[(12)\]

\[\text{Properties characterizing FRs}\]

\[\text{a. } [\text{cp wh- (COMP) } \ldots \ldots \text{DP}]_{\text{PP}}\]

\[\text{b. } [-D, -N, +WH]\]

The bracketed strings in (5) and (6) above are examples of FRs in English. The FR in (6) is introduced by the wh-word where, while the FR in (5) is introduced by the wh-phrase whatever painting. All three main varieties of FRs that are attested crosslinguistically occur in our languages as well: maximal free relative clauses, existential free relative clauses, and free-choice free relative clauses. The crucial distinction is semantic in nature, as revealed by the labels for the different kinds of FRs. This distinction often correlates with morpho-syntactic differences as well.\(^7\) We discuss and exemplify each variety of FR in turn in Sections 4.1–4.3.

\(^6\) Abbreviations used in all the tables. Language family names: UTO-AZT: Uto-Aztecan; MI-ZO: Mixe-Zoquean; CHI: Chibchan. Language names: Te: Southeastern Tepehuan; Na: Tlaxcala Nahuatl; AO: Acazulco Otomi; Ma: Matlatzinca; IM: Iliatenco Meꞌphaa; SZ: San Pedro Mixtepec Zapotec; Kꞌ: Kꞌiche; Qꞌ: Qꞌanjobꞌal; CJ: Chuj; Cl: Chꞌol; TT: Tsotsil and Tseltal; YM: Yucatec Maya; SP: Sierra Popoluca; Pe: Pesh.

\(^7\) See Simik (2017, forthcoming) for a thorough overview of the semantic properties of FRs and the analyses that have been suggested, and van Riemsdijk (2017) for a detailed overview of their syntactic properties and related syntactic proposals. Caioneiro et al (2013) investigate FRs in two Mixtec languages following a paradigm similar to ours.
4.1. Maximal free relative clauses

Maximal free relative clauses (Max-FRs) are those FRs that satisfy the properties in (13).

(13) Properties characterizing Max-FRs:

a. DEFINITENESS. A Max-FR can be replaced and paraphrased by a definite DP—a DP introduced by a definite marker or determiner in a language that has them, like the in English or el/la/los/las in Spanish—or by a PP with a definite DP as its complement.

b. REFERENTIALITY. A Max-FR is interpreted as referential: it refers to an individual. In this respect, Max-FRs are like proper names, definite DPs, and DPs introduced by demonstratives.

c. MAXIMALITY: A Max-FR is interpreted as maximal: it refers to the largest (‘maximal’) individual of a set of individuals. This is the same semantic behavior as seen with definite DPs.

Examples of Max-FRs are given in brackets in (14)a and (15)a with the wh-words introducing them in bold.

(14) a. Paloma revised \[\text{Max-FR} \text{what} \text{Frida wrote} \].
   b. Paloma revised \[\text{Definite DP} \text{the thing(s) Frida wrote} \].

(15) a. I went \[\text{Max-FR} \text{where} \text{my friends are vacationing} \].
   b. I went \[\text{PP to} \text{Definite DP} \text{the place(s) my friends are vacationing} \].

The Max-FRs in (14)a and (15)a satisfy the “Definiteness” property in (13)a: they can be replaced and paraphrased with definite DPs, as shown in (14)b and (15)b. They also satisfy the “Referentiality” property in (13)b: the Max-FR in (14)a refers to the object(s) Frida wrote and the Max-FRs in (15)a to the place(s) where the speaker’s friends are on vacation. Notice that referentiality is also a semantic property of the definite DPs that replaces the Max-FRs in (14)b and (15)b. Last, all the Max-FRs in (14)a and (15)a exhibit the “Maximality” property in (13)c. If Frida wrote a novel, a short story, and a poem, the Max-FR in (14)a refers to the plural individual that is formed by combining those three atomic individuals. Crucially, it cannot refer to anything smaller than that—such as the atomic individual consisting of the poem or the non-maximal plural individual made up of only the poem and the short story. This is the same semantic behavior as found with the plural definite DP \text{the writings by Frida}. Notice that the Max-FR in (14)a cannot be interpreted as triggering quantification over a set of individuals, unlike the indefinite DP \text{some of the things Frida wrote}. The same semantic behavior is observed when a set of places is involved: the Max-FR has to refer to the maximal individual of that set. For instance, the Max-FR in (15)a means the same thing as the PP \text{to the places where my friends are vacationing} with a definite DP as its complement. It cannot mean the same as the PP \text{to some of the places where my friends are vacationing} with an indefinite DP as its complement.
This semantic view of Max-FRs is based on Jacobson (1995), Dayal (1996), and Caponigro (2003, 2004), who, in turn, are inspired by the analysis of definite DPs in Sharvy (1980) and Link (1983). According to this approach, the Max-FR in (14)a receives the semantic derivation in (16).

(16) *Semantics derivation of the Max-FR in (14)a:*

\[
\begin{align*}
\text{Type-shifting} & \quad \text{CP}_2 \sim \lambda x_1[\text{non-human}(x_1) \wedge \text{wrote}(f, x_1)] \\
\text{CP}_1 \sim \lambda x_1[\text{non-human}(x_1) \wedge \text{wrote}(f, x_1)] \\
\\n\text{what}_1 & \sim \lambda p_{\text{ex},t} \lambda x_1[\text{non-human}(x_1) \wedge p(x_1)] \\
\text{C'} & \sim \lambda x_1[\text{wrote}(f, x_1)] \\
\text{IP} & \sim \text{wrote}(f, x_1) \\
\lambda_1 & \sim \text{Frida wrote } t_1
\end{align*}
\]

The IP of a Max-FR and, as we’ll see, of a FR in general, denotes an open proposition with the free variable \(x_1\) introduced by the \textit{wh}-trace. The familiar lambda abstraction over a \textit{wh}-variable occurs at the level of \(C'\), which ends up denoting a set of individuals—all and only the individuals that Frida wrote. The \textit{wh}-word \textit{what}, acting as a set restrictor, applies to the set of individuals that Frida wrote and returns its non-human subset—the denotation of \(\text{CP}_1\). A type mismatch now occurs. The Max-FR (\(\text{CP}_1\)) denotes a set of individuals, while the matrix predicate \textit{revise} selects for an individual-denoting object. A general meaning-preserving type-shifting operation is assumed to apply and turn the denotation of \(\text{CP}_1\) into the maximal individual of the set of non-human individuals that Frida wrote—the final denotation of the Max-FR (\(\text{CP}_2\)).

Max-FRs are extremely productive in all the Mesoamerican languages we have investigated and, in each language, they make use of most of the \textit{wh}-words, as shown in Table 2.

<p>| Table 2. Productivity of Max-FRs and their \textit{wh}-words in the languages under investigation |
|--------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|</p>
<table>
<thead>
<tr>
<th></th>
<th>UTO-AZT</th>
<th>OTO-MANGUEAN</th>
<th>MAYAN</th>
<th>M-Z</th>
<th>CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘who’</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>%</td>
</tr>
<tr>
<td>‘what’</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘where’</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘when’</td>
<td>√</td>
<td>*</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘why’</td>
<td>√</td>
<td>--</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘how’</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘what/which’ (+NP)</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>‘how much/many’ (+NP)</td>
<td>√</td>
<td>*</td>
<td>*</td>
<td>√</td>
<td>?</td>
</tr>
</tbody>
</table>

\textbf{Note: Marks:} √: attested; *: not attested; ?: unclear if attested; --: no simple \textit{wh}-word conveying the relevant meaning; %: speaker variation.
Pesh is the only language with a non-productive system of Max-FRs: only the wh-word for ‘when’ introduces Max-FRs. It may be by chance, but Pesh is also the only non-Mesoamerican language among the languages we have investigated.

The shaded two columns in Table 2 highlight the fact that K’iche’ and San Pedro Mixtepec Zapotec enrich the wh-words introducing Max-FRs with extra morpho-syntactic marking, like Modern Greek and Hungarian. K’iche’ adds what looks identical to a definite determiner, le, right after the wh-word, as shown in bold in (17).

(17) X-ki-muli-j k-iib’ [ jachin taq le k-e-xajow-ik].\(^8\) K’iche’
COMPL-A1PL-gather-ACT A1PL-RECPwho PL DET ICP-B3PL-dance-SS
‘The ones who dance, gathered.’

Definite determiners must precede nominals in K’iche’. Can Pixabaj (2020) leaves it open whether the definite D occurring with wh-words in Max-FRs is a syntactically independent D head or a suffix combing with the wh-root (and other suffixes). San Pedro Mixtepec Zapotec, instead, makes use of what is clearly a prefix, tèl-, to characterize the wh-words in Max-FRs, as shown in (18). This prefix doesn’t resemble any other marker in the language.

(18) d-áw ná [ tél-pè b- dziél]\(^9\) San Pedro Mixtepec Zapotec
COMPL-eat 1SG TEL-WH.INAN COMPL-find
‘I ate what was found.’

These facts provide indirect support to the type-shifting rule that we have assumed in the semantic derivation in (16). The meaning of a Max-FR has to shift from a set to its maximal individual: this meaning shift is triggered either lexically or by general type-shifting rules that deal with type mismatches. The first option requires the language to have an overt lexical item/morpheme that (i) conveys the correct type-shifting operator and (ii) can morpho-syntactically combine with the whole FR (a CP) or its wh-word. K’iche’ and San Pedro Mixtepec Zapotec have such a lexical item or morpheme. The meaning of wh-words in Max-FRs in these two languages looks like the one schematized in (19): a function that applies to a set of individuals to return the maximal individual of a subset that satisfies the “WH” restriction (human, non-human, location, etc.). The shift from a set to an individual is now triggered by a lexically encoded \(\mathsf{\iota}\) operator. For comparison, the meaning of the wh-words in Max-FRs in all the other languages we have investigated can be represented as in (20).

(19) ‘the’+wh \(\leadsto\) \(\lambda x_1[WH(x_1) \land P(x_1)]\)

(20) wh \(\leadsto\) \(\lambda x_1[WH(x_1) \land P(x_1)]\)

\(^8\) From Can Pixabaj (2020: ex. 40). The abbreviations in the glosses of this example and all the others follow “The Leipzig Glossing Rules—List of Standard Abbreviations, Updated on May 31, 2015”. We refer to the source of the examples for the explanation of further abbreviations that individual authors may have used.

The option in (19) is the least common among our languages; the default seems to be the free application of the appropriate type shifter without overt lexical marking. This fact mirrors and supports previous findings about Max-FRs in non-Mesoamerican languages.

The wh-word for ‘why’ and the wh-phrase ‘what’/‘which’+NP can introduce Max-FRs in several of our languages. This possibility is rarer in Max-FRs in Indo-European, although it has been recently documented for Romanian (Caponigro 2019, Caponigro & Fălăuş in preparation). Therefore, any successful attempt to explain this restriction in Indo-European cannot be derived from non-gradable/non-parametrizable grammatical principles (contra Donati & Checchetto 2011 a.o.).

Last, we have found Max-FRs in the argument and adjunct positions where DPs or PPs can occur, without any particular restriction. Once again, this mirrors previous findings about Max-FRs in other languages. This distributional freedom is a crucial syntactic difference between Max-FRs and the kind of FR we discuss next.

4.2. Existential free relative clauses

Existential free relative clauses (henceforth, Ex-FRs) are those FRs satisfying the properties in (21)a–b and exhibiting the syntactic structure schematized in (21)c.\(^\text{10}\)

\[(21)\] Properties characterizing Ex-FRs:

a. EXISTENTIAL MEANING. Ex-FRs can be replaced and paraphrased by existentially quantified nominal expressions—indefinite DPs, which are introduced by indefinite markers (e.g., determiners like a in English or un/una/unos/unas in Spanish), or bare nominals (e.g., I love books), in languages that allow for either.

b. EXISTENTIAL PREDICATE. When attested in a language, Ex-FRs can always occur as the complement of existential ‘be’ and existential ‘have’ in that language.

c. \[\text{[V}_{\text{existential]} \ [\text{CP wh- (REL/COMP)} \ldots \ldots \ldots \text{Indefinite DP/Bare Nominal} ]}\]

Ex-FRs are not attested in English or other Germanic languages,\(^\text{11}\) but are common in Romance, Balto-Slavic, and Semitic languages, as well as our languages. Examples of Ex-FRs from Italian are given in brackets in (22)a–(24)a.

\[(22)\] a. C’è [Ex-FR \textbf{chi} dice sempre sì].
   ‘There are people who say yes all the time.’

---

\(\text{10}\) See Izvorski (1998), Grosu (2004, 2013), Caponigro (2003, 2004), Šimík (2011), and work cited there. Šimík (2017) provides a thorough review of the relevant literature on Ex-FRs and related constructions. We assume that they are clauses (CPs). Grosu (2004, 2013) and Šimík (2011) argue that Ex-FRs should not even be called FRs because they are not necessarily clauses. They propose the label Modal Existential Construction (MEC), instead.

\(\text{11}\) With the possible exception of Yiddish and Yiddish-influenced varieties of English (Caponigro 2003: §3.2.4).
b. Ci sono [Indefinite DP (delle) persone che dicono sempre sì].

‘There are people who say yes all the time.’

(23) a. Non aveva [Ex-FR chi le curasse i bambini].

‘She didn’t have anybody who looked after her children.’

b. Non aveva [Indefinite DP nessuno che le curasse i bambini].

‘She didn’t have anybody who looked after her children.’

(24) a. Ho [Ex-FR di che lamentarmi].

‘I have something/things to complain about.’

b. Ho [Indefinite DP (delle) cose di cui lamentarmi].

‘I have things to complain about.’

All the Ex-FRs in (22)a–(24)a satisfy the properties in (21)a–b. Examples (22)b–(24)b show that the Ex-FRs can be replaced and paraphrased with existentially quantified DPs, as required by the “Existential Meaning” property in (21)a. For instance, (22)a asserts the existence of at least one person who says yes all the time. In other words, the set of the people who say yes all the time is non-empty. The Ex-FR is, therefore, semantically close to the indefinite DP in (22)b. Example (23)a with negation on the matrix existential predicate asserts that it is not true that there is at least one individual in the set of people who looked after the children of the contextually salient woman that the speaker is referring to. In other words, that woman didn’t have anybody who looked after her children. Therefore, the Ex-FR in (23)a is semantically close to the negative indefinite in (23)b. As for the “Existential Predicate” property in (21)b, all the Ex-FRs in (22)a–(24) occur as the complements of existential ‘be’ or existential ‘have’ in Italian. The two properties are, of course, connected. Existential predicates select for indefinites as their complements (exclusively so in many languages). It is not surprising, therefore, that if wh-clauses are allowed as their complements, they will be interpreted as indefinite DPs.

Although with significant variants, all the semantics analyses that have been suggested for Ex-FRs agree that the existential force of an Ex-FR is crucially related to the matrix predicate. In (25), we sketch the semantic derivation of the Ex-FR in (22)a, following Caponigro (2003, 2004).

The semantics derivation of an Ex-FR is identical to the one of a Max-FR up to CP (CP₁ in a Max-FR): the CPs in both FRs denote a set of individuals. The crucial difference is the assumption that there’s no type mismatch between an Ex-FR and its matrix predicate. An existential predicate (there’s in (25)) selects for a set-denoting complement and existentially quantifies over it, as shown by the logical translation of the top-most matrix IP in (25).
Ex-FRs are extremely productive in all the languages we have investigated and, in each language, they make use of most of the \textit{wh}-words. Table 3 summarizes our findings.

\textbf{Table 3. Productivity of Ex-FRs and their \textit{wh}-words in the languages under investigation}

<table>
<thead>
<tr>
<th>\multicolumn{1}{c}{\textit{who}}</th>
<th>\multicolumn{1}{c}{\textit{what}}</th>
<th>\multicolumn{1}{c}{\textit{where}}</th>
<th>\multicolumn{1}{c}{\textit{when}}</th>
<th>\multicolumn{1}{c}{\textit{why}}</th>
<th>\multicolumn{1}{c}{\textit{how}}</th>
<th>\multicolumn{1}{c}{\textit{what/which}}</th>
<th>\multicolumn{1}{c}{\textit{how much/many}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTO-AZT</td>
<td>Te</td>
<td>Na</td>
<td>AO</td>
<td>Ma</td>
<td>IM</td>
<td>SZ</td>
<td>MAYAN</td>
</tr>
<tr>
<td>`who'</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>`what'</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>`where'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>`when'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>`why'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>`how'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>`what/which'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+NP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>`how much/many'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textbf{Note: Marks:} √: attested; *: not attested; ?: unclear if attested; --: no simple \textit{wh}-word conveying the relevant meaning; %: speaker variation.

The shaded columns in Table 3 highlight the four languages for which we found mood restrictions in their Ex-FRs. In those languages, Ex-FRs can only occur with an irrealis marker. This restriction correlates with Šimik’s (2011) findings that most of the languages he surveyed require the predicate of an Ex-FR to be in the infinitive or subjunctive form. On the other hand, at least one Indo-European language is found—Italian—that allows for indicative, subjunctive, or infinitive in its Ex-FRs, as shown in (22)a–(24)a. Crucially, twelve languages we have investigated don’t exhibit any TAM (Tense/Aspect/Mood) restriction in their Ex-FRs either, bringing further

\footnote{Šimik’s (2011) mainly surveys different groups of Indo-European languages, but also includes some Finno-Ugrian and Semitic languages.}
evidence that the pattern previously found in Italian is not an anomaly. For instance, the predicates of both bracketed Ex-FRs in the example in (26) from San Pedro Mixtepec Zapotec and the one in (27) from Yucatec Maya carry a complettive aspect affix (in bold) without any modal component.

(26) y-ô [ pè p-kà Màří].\textsuperscript{13} \hspace{1cm} \textit{San Pedro Mixtepec Zapotec}
\begin{tabular}{llll}
\textsc{stat-exist} & \textsc{wh.inan} & \textsc{compl-buy} & Màří \\
\end{tabular}

‘There is something that María bought.’

(27) Yaan [ ba'axten t-u meet-aj]. \textsuperscript{14} \hspace{1cm} \textit{Yucatec Maya}
\begin{tabular}{llll}
\textsc{ex} & \textsc{why} & \textsc{pfv-a3} & \textsc{do.tr-ss.cplv} \\
\end{tabular}

‘There is a reason why he did it.’

4.3. Free-choice free relative clauses

Free-choice free relative clauses (henceforth, FC-FRs) are those FRs that satisfy the properties in (28)a–b.\textsuperscript{15} Examples are given in (29) and (30).

(28) \textit{Properties characterizing FC-FRs:}
\begin{enumerate}
\item \textsc{free-choice inference}. A sentence containing an FC-FR obligatorily triggers an inference of ignorance or indifference.
\item \textsc{free-choice marker}. An FC-FR always contain a free-choice (FC) marker.
\end{enumerate}

(29) a. [\textsc{fc-fr whatever} \text{Paloma is cooking right now}] uses onions.\textsuperscript{16}
\hspace{1cm} \textit{asserted content}: [\textsc{max-fr/dp}\{\text{What}\}/\{\text{The stuff that}\} \text{Paloma is cooking right now}] uses onions.
\hspace{1cm} \textit{ignorance FC inference}: The speaker doesn’t know what Paloma is cooking right now.

(30) a. Pablo (simply) voted for [\textsc{fc-fr whoever} \text{was at the top of the ballot}].\textsuperscript{17}
\hspace{1cm} \textit{asserted content}: Pablo voted for [\textsc{dp} \text{the person who was at the top of the ballot}].
\hspace{1cm} \textit{indifference FC inference}: Pablo didn’t care about who was at the top of the ballot.

Example (29)a shows a bracketed FC-FR in the subject position of its matrix clause. It is introduced by the bolded \textit{wh}-word \textit{whatever}, which results from the morphological enrichment of the \textit{wh}-root \textit{what} with the FC suffix \textit{-ever}. Following Dayal’s (1997) seminal analysis for English and von Fintel’s (2000) further development, we can look at the meaning of the FC-FR in (29)a

\begin{itemize}
\item \textsuperscript{13} From Antonio-Ramos (2020: ex. 55).
\item \textsuperscript{14} From AnderBois & Chan Dzul (2020: ex. 23c).
\item \textsuperscript{15} See Dayal (1997) and von Fintel (2000). Also, see Šimik (forthcoming) for an overview of the relevant issues and literature. Free-choice free relative clauses in English (and in other languages as well) are often labeled “\textit{-ever} free relative clauses”.
\item \textsuperscript{16} Adapted from Dayal (1997: ex. 27a).
\item \textsuperscript{17} Adapted from von Fintel (2000: ex. 18).
\end{itemize}
as contributing at two different levels. The asserted content of (29)a is the same as the one of (29)b, in which the FC-FR has been replaced with a Max-FR or a definite DP. The FC-FR in (29)a, though, obligatorily triggers the ignorance inference (a presupposition, according to von Fintel) that the speaker doesn’t know the identity of what Paloma is cooking, as stated in (29)c. Notice that both the Max-FR and the definite DP in (29)b are compatible with a situation in which the speaker doesn’t know the identity of what Paloma is cooking, but, crucially, they are also compatible with a situation in which the speaker is not ignorant about what Paloma is cooking. In other words, they do not obligatorily trigger an ignorance inference.

Example (30)a shows a bracketed FC-FR in the complement position of the preposition for in the matrix clause. It is introduced by the morphologically enriched wh-word whoever in bold. Example (30)a asserts the same as (30)b, in which the FC-FR has been replaced with a definite DP. Unlike (30)b, though, (30)a with an FC-FR necessarily triggers the indifference inference that Pablo—the individual the matrix subject refers to—doesn’t care about the actual identity of the candidate who was at the top of the ballot, as stated in (30)c.

We group these two inferences under the same “free choice” label to highlight the fact that they both trigger a form of variation on the reference of an FC-FR. It does refer to a maximal individual, like a Max-FR, but which maximal individual it refers to can vary depending on the relevant modality and the related modal agent: epistemic modality and the speaker in (29), or counterfactual modality and Pablo—the individual the matrix subject refers to—in (30).

It is a parameter of crosslinguistic variation whether both inferences are triggered in FC-FRs or only one—and, if only one, which one. English FC-FRs allow for either, depending on various factors (e.g., tense, aspect, discourse conditions). Italian and Romanian FC-FRs allow only for ignorance inferences, and not exactly of the same kind as those that are triggered by FC-FRs in English. Caponigro & Fălăuş 2018 argue for an analysis of FC-FRs in Italian and Romanian that is significantly different from Dayal’s and von Fintel’s since it analogizes FC-FRs to indefinite DPs introduced by FC any in English, rather than definite DPs.

FC-FRs are attested in all the languages we have investigated but Southeastern Tepehuan (O’dam), Tsotsil, and Tseltal, as summarized in Table 4. Further investigation is needed to understand whether Ch’ol has true FC-FRs.
Table 4. Productivity of FC-FRs and their wh-words in the languages under investigation

<table>
<thead>
<tr>
<th></th>
<th>UTO-AZT</th>
<th>OTO-MANGUEAN</th>
<th>MAYAN</th>
<th>M-Z</th>
<th>CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Te</td>
<td>Na</td>
<td>AO</td>
<td>Ma</td>
<td>IM</td>
</tr>
<tr>
<td>'who'</td>
<td>*</td>
<td>√</td>
<td>√</td>
<td>*</td>
<td>√</td>
</tr>
<tr>
<td>'what'</td>
<td>*</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>'where'</td>
<td>*</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>'when'</td>
<td>*</td>
<td>√</td>
<td>*</td>
<td>*</td>
<td>√</td>
</tr>
<tr>
<td>'why'</td>
<td>*</td>
<td>--</td>
<td>*</td>
<td>*</td>
<td>√</td>
</tr>
<tr>
<td>'which/what' (+NP)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>'how much/many' (+NP)</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>?</td>
<td>√</td>
</tr>
</tbody>
</table>

**NOTE: Marks:** √: attested; *: not attested; ?: unclear if attested; --: no simple wh-word conveying the relevant meaning; n/a: data not available; F+: free choice markers precede wh-words; +F: free choice markers follow wh-words; FF: free choice marking results from reduplication of wh-expressions.

The languages with FC-FRs mark them by means of FC markers preceding (F+) or following (+F) wh-words, as shown in (31) with the bolded FC suffix in Tlaxcala Náhuatl.

(31) Ø-miki-s [ak-sa i-nawa-k ti-mo-namik-ti-s]18
    S3-die-IRR who-FC POSS3SG-ear-LOC S2SG-RR-find-CAUS-IRR

    ‘Whoever you get married to will die.’

Only Pesh uses reduplication of wh-words to mark FC, as shown in (32).

(32) í’iri tàkàtùhùmà ápáhá19
    [í=ra i=ra ta-ka-tuh-u-wa=ma]
    DEM.PROX=ABS DEM.PROX=ABS OBJ1-APPL.R-cook-SBJ2-PRS=CRT
    Ø-à-pa=hà?
    OBJ3SG-eat- SBJ1SG.FUT=FOC

    ‘I will eat whatever you cook for me.’

Further research is needed to establish which inferences FC-FRs trigger in each language and their exact nature. What is crucial for our assessment of the syntax-semantics mapping generalization in (1) is that neither kind of inference is obligatorily triggered without overt morpho-syntactic marking: a FC marker of some kind is needed.

---

18 From Flores-Najera (2020: ex. 50).
19 From Chamoreau (2020: ex. 58).
This concludes our introduction to the three main varieties of FRs that are attested across languages and those of their features that have been particularly relevant for our investigation. In the next section, we discuss a different kind of [−H]RC.

5. Light-headed relative clauses

Light-headed relative clauses (henceforth, LHRs) are [−H]RCs with an overt D head that can co-occur with a wh-expression, a relative marker, a complementizer, or no marker at all. The characterizing features of LHRs are summarized by the syntactic schema in (33)a (inspired by the syntactic analysis in Citko 2004) and the feature bundle in (33)b.

\[(33) \text{Properties characterizing LHRs:}\]
\[a. \ [D \ [CP (\text{wh-/REL/COMP}) \ldots \ldots ] ]_{DP/PP}\]
\[b. \ [+D, -N, \pm WH]\]

D heads in LHRs can be of three main kinds, though they are not necessarily all instantiated in all languages with LHRs: articles, demonstratives, or quantifiers. Some D heads behave like pronominals and never take an NP complement; some others only allow an adnominal use with an NP complement when they are not introducing an LHR; and some exhibit both patterns. What all D heads share is that they occur—with or without other material—where DPs (or PPs) can occur and none of them has N features (e.g., none of them can occur as the complement of a D or can refer to a set of individuals). Examples of LHRs are given in (34)–(37), with D heads in bold and the whole LHRs in brackets. Comments follow.

\[(34) \text{Article D + COMP}\]
\[[\text{LHR Lo que tú crees}] \text{ no es cierto.}^{21}\] \[\text{Spanish}\]
\[\text{the.N COMP you.SG believe.2SG not is true}\]
\‘{What}/[That which} you believe is not true.’

\[(35) \text{Demonstrative D + WH/COMP/∅}\]
\[\text{I like [LHR those who/that/∅ you like].}\]

\[(36) \text{Universal D + ∅}\]
\[[\text{LHR Tuláakal k-u púuts'-ul-o'ob}, k-u y-áalkab-∅-o'ob}^{22}\] \[\text{Yucatec Maya}\]
\[\text{all IPFV-A3 flee-SS.ICPLV-B3PL IPFV-A3 EP-run-SS.ICPLV-B3PL}\]
\‘All those who escape run (from him).’

---

\[20\text{ We borrow the label “light-headed relative clauses” from Citko (2004), although we expand its coverage by also including constructions that lack wh-expressions.}\]
\[21\text{ From Plann (1980: ex. I.1.a). Glosses and translation have been adapted.}\]
\[22\text{ From AnderBois & Chan Dzul (2020: ex. 33b).}\]
Example (34) shows an LHR from Spanish that is introduced by a definite article followed by a declarative complementizer. Example (35) shows an LHR in English that is introduced by a demonstrative D head and can be followed by a wh-word (and/or relative pronoun), a declarative complementizer, or no marker at all. Example (36) shows that Yucatec Maya can use a universal quantifier as the D head of its LHRs, while example (37) shows that an indefinite D can introduce LHRs in Iliatenco Me’phaa.

Overall, LHRs are extremely productive across the languages we have investigated, as summarized in Table 5.

### Table 5. Productivity of LHRs in the languages under investigation

<table>
<thead>
<tr>
<th>LHR</th>
<th>UTO-AZT</th>
<th>OTO-MANGUEAN</th>
<th>MAYAN</th>
<th>MI-ZO</th>
<th>CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Te Na AO Ma IM SZ K' Q' Cj Cl TT YM SP Pe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LHR</td>
<td>√ √ √ √ √ √ √ √ √ √ √ √ √ √ √</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Mark: √: attested.

The picture that emerges from our investigation depicts LHRs as a family of constructions, rather than just a single construction. Across languages and even within the same language, there may be LHRs whose morpho-syntax is closer to that of [−H]RCs or headed relative clauses or neither. The details about our findings can be found in Caponigro et al. (2020), including the data about which kind of D head can introduce an LHR in which language.

One of the main goals of this paper is to test the syntax-semantics mapping generalization in (1). Therefore, the semantic behavior of LHRs is crucial for us. Our findings show that it is the D of an LHR that is the element that determines the kind of meaning that the LHR conveys. If the D head is a definite or demonstrative D, then the whole LHR behaves like a referential DP. If the D head is a quantifier, then the LHR behaves like a quantificational DP with the same quantificational strength as its D head.

We are not aware of any compositional semantic analysis of LHRs. We tentatively sketch one in (38) and (39) that accounts for our empirical finings and our generalization. The analysis in (38) applies to LHRs that have an overt lexical item licensing a wh- trace/variable, may it be a wh-

---

23 From Duncan & Torrence (2020: ex. 64b).
24 See Plann (1980: Ch. 1) for arguments that lo is a definite article rather than a pronoun. The article lo is never used with an NP complement in Spanish.
25 While presenting this paper at SULA 11, Scott AnderBois (p.c.) informed me that he had started working on a compositional semantics of LHRs and had presented some preliminary results in a lecture at CIESAS on July 13, 2020.
word or a non-\textit{wh} relative pronoun. For simplicity, we refer to either in the derivation in (38) by means of the label \textit{wh}_1.

(38) \textit{Semantics derivation of LHRs with a \textit{wh}-word or a relative pronoun:}

\[
\begin{array}{c}
\text{DP} \rightarrow \lambda x_1 [\text{WH}(x_1) \land \text{V}(x_1)] \\
\text{CP} \rightarrow \lambda x_1 \left[\text{WH}(x_1) \land \text{P}(x_1)\right] \\
\text{wh}_1 \rightarrow \lambda P_{x_1} \lambda x_1 [\text{WH}(x_1) \land P(x_1)] \\
\text{C'} \rightarrow \lambda x_1 [\text{V}(x_1) \ldots] \\
\text{C} \rightarrow \lambda_1 \text{IP} \rightarrow [\text{V}(x_1) \ldots] \\
\ldots t_1 \ldots
\end{array}
\]

The semantic derivation of an LHR is the same as the semantic derivation of a FR up to CP: a subset of individuals resulting from the restriction (WH) that the \textit{wh}-word or relative pronoun \textit{wh}_1 imposes on the set of individuals that is denoted by \textit{C'} after abstracting over the \textit{wh}-variable licensed by \textit{wh}_1 (V stands for whatever 1-place predicate results from the semantic derivation of the LHR up to its IP). The D head is looking for a set denoting complement to return either an individual, if D is a definite article or a demonstrative, or a generalized quantifier, if D is quantificational.

The semantic derivation in (39), instead, applies to LHRs that lack a \textit{wh}-word or a relative pronoun. It’s a simplified version of (38) in which the denotation of \textit{C'} is simply passed up to its mother node CP because of the lack of any restriction coming from a \textit{wh}-word or a relative pronoun.

(39) \textit{Semantics derivation of LHRs without a \textit{wh}-word or a relative pronoun:}

\[
\begin{array}{c}
\text{DP} \rightarrow \lambda x_1 [\text{V}(x_1) \ldots] \\
\text{CP} \rightarrow \lambda x_1 [\text{V}(x_1) \ldots] \\
\text{C'} \rightarrow \lambda x_1 [\text{V}(x_1) \ldots] \\
\text{C} \rightarrow \lambda_1 \text{IP} \rightarrow [\text{V}(x_1) \ldots] \\
\ldots t_1 \ldots
\end{array}
\]
6. Super-free relative clauses

Super-free relative clauses (henceforth, SFRs) are the last variety of [−H]RCs that we have investigated. Like all [−H]RCs, they lack an N head. Unlike LHRs, they lack a D head as well, resembling FRs in this regard. But they are even “freer” than FRs in lacking a wh-expression as well. This is why we have labelled them “super-free”. The syntactic structure we have assumed for SFRs is summarized in (40)a and their main features are bundled in (40)b.

(40) Properties characterizing SFRs:
   a. \[\text{CP (REL/COMP) } \ldots \ldots \text{DP/PP}\]
   b. \[\text{[−D, −N, −WH]}\]

SFRs have been largely ignored in the literature. We are not aware of any detailed description of them, let alone a systematic or crosslinguistic investigation. Our study is the first one to precisely define and investigate SFRs. Our findings show that SFRs are well-attested in our languages, as summarized in Table 6.

Table 6. Productivity of SFRs in the languages under investigation

<table>
<thead>
<tr>
<th></th>
<th>UTO-AZT</th>
<th>OTO-MANGUEAN</th>
<th>MAYAN</th>
<th>MI-ZO</th>
<th>CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Te</td>
<td>Na</td>
<td>AO</td>
<td>Ma</td>
<td>IM</td>
</tr>
<tr>
<td>SFR</td>
<td>*</td>
<td>√</td>
<td>*</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

**NOTE:** Marks: √: attested; *: not attested; ?: unclear if attested.

Out of our fifteen languages, eight have SFRs for sure, while two don’t. The remaining five languages (those with a question mark in Table 6) have constructions that look like SFRs, but the scholars who have investigated them in those five languages leave the issue open whether they are true [−H]RCs or headed relative clauses with a silent nominal head (see the relevant language-specific chapters in Caponigro et al. 2020 for the language-specific data and discussions).

Morpho-syntactically, SFRs can be introduced by a complementizer, as in (41), or by a non-wh relative pronoun, as in (42), or by no marker at all, as in (43) and (44). Crucially, the bracketed SFRs in (41)–(44) are full, tensed clauses, rather than nominalized tenseless clauses or subclausal constituents.

(41) COMP / REFERENTIAL
   \text{y=O-Ø-asi-ko} \quad [\text{SFR den} \quad \text{i-pampa} \quad \text{o-ti-choka-ya}]^{27} \quad \text{Tlaxcala Náhuatl}
   \text{already=PST-S3-arrive-VEN.PST} \quad \text{COMP} \quad \text{POSS3SG-RSN} \quad \text{PST-S2SG-cry-IPFV}
   \text{‘The one you cried for has already arrived.’}

---

26 As a reminder, “TT” stands for Tsotsil and Tzeltal, two closely related but different languages.
27 Adapted from Flores-Nájera (2020: ex. 66).
(42) REL / EXISTENTIAL
   \emptyset = \text{it-wi} \quad [\text{SFR} \ _\text{ta=ku?a?m-ket-ne?-wi=pV?}]^{28} \quad \text{Sierra Popoluca}
   3\text{ABS} = \text{be-COMPL} \quad 1\text{ABS} : \text{INCL} = \text{search-descend-PERF-COMPL} = \text{REL}
   ‘There is someone who looks after us.’

(43) \emptyset / EXISTENTIAL
   Ay-\emptyset \quad [\text{SFR} \ _\text{max-\emptyset \ w-aq’-ok-toq \ y-ul \ refri}]^{29} \quad \text{Q’anjob’al}
   \text{EXS-B3} \quad \text{COM-B3} \quad 1\text{ASG} : \text{give-DIR-DIR} \quad 3\text{in} \ \text{fridge}
   ‘There is some [chicken] that I put into the fridge.’

(44) \emptyset / REFERENTIAL
   \text{no \ tata, \ cha’ \ phi} \quad [\text{SFR} \ _\text{gahch’ówi \ chhan \ tu \ hëwi \ _}]^{30} \quad \text{Matlazinca}
   \text{no \ sir} \quad \text{NEG} \quad \text{be.so} \quad \text{2PLPRO} \quad \text{2PL.AMBU} \quad \text{INFL} \quad \text{do}
   ‘No sir, what you (pl) are doing is not good.’

Semantically, SFRs across our languages are interpreted as referential and maximal by default, as shown in (41) and (44). If a language allows SFRs to occur as the complement of existential predicates, then they are interpreted as indefinites as well, as shown in (42) and (43). There are languages that only allow SFRs as the complement of existential predicates: all our Mayan languages but Ch’ol behave this way, as shown in (43) for Q’anjob’al.

This consistent semantic behavior of the same construction across languages and language families cannot be by chance and fully supports our generalization in (1). We propose to account for it as in (45).

(45) Semantic derivations of SFRs:

Option 2 - Existential, like Ex-FRs
\quad \text{VP} \rightarrow \exists x_1 \ldots
\quad \text{Exist. V} \rightarrow \lambda P \exists xP(x)

Option 1 - Referential, like Max-FRs
\quad \text{CP}_2 \rightarrow t x_1 \ldots
\quad \text{Type-shifting}
\quad \text{CP}_1 \rightarrow \lambda x_1 [V(x_1)]
\quad \text{C'} \rightarrow \lambda x_1 [V(x_1) \ldots]
\quad \text{C} \quad \text{IP} \rightarrow [V(x_1) \ldots]
\quad \lambda_1 \quad \ldots \ t_1 \ldots

\text{---}

28 Adapted from López Márquez (2020: ex. 89b).
29 Adapted from Mateo Toledo (2020: ex. 67).
30 Adapted from Palancar & Carranza Martínez (2020: ex. 58).
The two semantics derivations in (45) share the same steps all the way up to CP1. This is also the very same semantic derivation as the one of LHRs without wh-words we proposed in (39). CP1 denotes a set of individuals in all these constructions. If an SFR occurs in a typical argument position in which an individual denoting constituent is required, then a type mismatch occurs. It is the same type mismatch as the one we already discussed for Max-FRs. Therefore, we adopt the same strategy to handle it: a type-shifting rule applies and turns the set denoted by CP1 into its maximal individual, as shown by Option 1 in (45). On the other hand, if an SFR occurs in the complement position of existential predicates like those we discussed for Ex-FRs, then the very same semantic process we proposed for Ex-FRs applies to these SFRs as well: the matrix predicate existentially quantifies over its set-denoting complement, as shown by Option 2 in (45).

7. Some general remarks and conclusions

We started with proposing the generalization about the syntax-semantics mapping in (1), according to which the interpretative options for [−H]RCs are highly restricted. We have supported this generalization with the findings of our novel crosslinguistic investigation of [−H]RCs in fifteen languages (all Mesoamerican but one) from five language families (see Caponigro et al. 2020 for the detailed findings). We have also provided precise definitions and semantic analyses for each of the main kinds of [−H]RCs that we have identified. Our proposals for the three different kinds of FRs build on definitions and analyses that have been proposed for mainly Indo-European languages. Our proposals for LHRs and SFRs, instead, are the first general definitions and semantic analyses of those kinds of [−H]RCs that we are aware of. All the proposals rely on a small set of shared assumptions that have been independently argued to be needed in the grammar.

Our generalization and analyses make clear predictions that we bring to the attention of the linguistic community for further testing: no language should be found in which [−H]RCs without an overt quantificational marker can be interpreted as quantificational expressions.31

Before concluding, we touch on some broader issues and further consequences of our findings. First, we would like to speculate on the reasons behind the generalization in (1). In other words, why do natural languages behave in the way that is merely described by the generalization in (1)? The semantic analyses that we have provided for the different kinds of [−H]RCs account for the pattern in (1), but don’t explain why [−H]RCs semantically behave in this way rather than in any other way. For instance, why can’t a [−H]RC freely denote any generalized quantifier, depending on the contexts? Why can’t the bracketed Max-FR in *Frida ate [what Pablo cooked]* receive the same interpretation as *Frida ate something that Pablo cooked* when uttered in a context in which it is made clear that out of Pablo’s three dishes, Frida ate only one? Or why can’t it mean *Frida

---

31 Gitksan (Tsimshianic) as presented in Aonuki (2020) may look like a possible counterexample. It is claimed that FRs without any special marking can be interpreted as indefinites in argument position freely, not just as complements of existential predicates. On the other hand, it has emerged from a follow-up Zoom conversation with Yurika Aonuki, that the wh-words that introduce FRs in Gitksan can also be used on their own as indefinites (e.g., the wh-word for ‘who’ can mean ‘someone’ when used on its own). It follows that the alleged indefinite FRs could just be headed (or light-headed) relative clauses with the indefinite wh-word as their head. Further investigation is needed to determine which hypothesis is correct.
ate most things that Pablo cooked in a context in which Pablo prepared ten things and Frida ate only six of them?

The core idea we would like to elaborate on is that some semantic operations are “information preserving”, while others are not. Consider the type-shifting rule we have assumed to account for the semantic behavior of Max-FRs and SFRs. It is information preserving. It turns a set of individuals into its maximal individual. This is just a different way of presenting the same piece of information. In fact, given a set, one can always build its maximal individual, and, given the maximal individual of a set, one can always retrieve the set from which that maximal individual originates. For instance, the set of the three sisters Olga (O), Masha (M), and Irina (I) in Anton Chekhov’s play can be represented as in (46)a, while its maximal individual can be represented as in (46)b: the plural individual resulting from the sum of all the atomic individuals in the set in (46)a. The one-to-one (injective) mapping between a set and its maximal individual can be made explicit by means of a join semi-lattice representation, as in (46)c: the set in (46)a contains all the atomic elements of the join semi-lattice, while the maximal individual in (46)b is the supremum or join of the join semi-lattice, as shown by the bolded top element in join semi-lattice in (46)c.

(46) a. *Set:* \{O,M,I\}
   b. *Maximal individual of \{O,M,I\}:* O⊕M⊕I
   c. *Join semi-lattice:*

   ![Join semi-lattice diagram]

Intuitively, if you have the set of Olga, Masha, and Irina in mind, then you can as easily think of the plural individual made of the three of them—“the three sisters.” The other way around holds as well: if you are thinking of the maximal individual “the three sisters”, you straightforwardly switch to thinking of the set containing Olga, Masha, and Irina without any additional information or instruction.

Quantification, instead, is not an information-preserving semantic operation. Turning a set of individuals into a generalized quantifier, i.e., a function from sets of individuals to truth values, is not a one-to-one (injective) mapping: many different quantifiers can be built out of the same set of individuals or domain restrictor. Therefore, once the semantic derivation of a [−H]RC reaches the set-denotation level, it would be undetermined which quantifier should be built out of that set, unless extra information is provided by means of overt lexical material. This would explain why [−H]RCs behave like quantificational DPs only (i) as the complement of existential predicates (Ex-FRs and some SFRs), or (ii) with overt D heads (LHRs), or (iii) with a *wh*-word carrying appropriate extra morphological marking (FC-FRs).
The previous considerations lead us to the next broader question: whether the generalization in (1) can be subsumed under a broader generalization concerning all natural language expressions that refer to individuals or quantify over individuals without overt markers of those semantic operations. The obvious other candidates to consider would be simple Ns without a D in an argument position, i.e., (singular and/or plural) bare nominals. In languages that allow them, bare nominals can be used to refer to maximal individuals or kinds, but they cannot be freely assigned quantificational force. In this respect, bare nominals behave like [−H]RCs in following the first part of the generalizations in (1). Unlike [−H]RCs, though, bare nominals can be interpreted as indefinites in a variety of argument positions, depending on the aspectual properties of their clause. For instance, the bolded English bare plural in the generic sentence in (47)a refers to a kind, while the one in the episodic sentence in (47)b receives a quantificational interpretation that makes it resemble a quantificational DP like some Italians

(47) a. **Italians** gesture a lot.
    b. **Italians** are yelling in the street.

Still, Italians cannot be interpreted as most Italians, half of the Italians, no Italian, or any other quantificational DP, unless it occurs with an overt quantificational element (e.g., a quantificational D or an adverb of quantification). In conclusion, [−H]RCs and bare nominals exhibit semantic similarities, but also important differences that we leave to future investigation.

Let’s conclude by briefly touching on further consequences of our investigation. We have revealed an articulated and productive world of [−H]RCs that has been largely ignored so far. An extensive typological database like Dryer & Haspelmath (2013) mentions [−H]RCs only to exclude them from their inventory of constructions. The study of contact–induced grammaticalization in Heine & Kuteva (2006: 204–205) starts from Haspelmath’s (1998: 279–280) observation that relative clauses that are introduced by *wh*-expressions are common in “Standard Average European” (SAE) and goes on to claim that they are much less common in other languages families and that, whenever they are attested outside SAE, this is due to language contact with SAE. Our findings clearly show that [−H]RCs that are introduced by *wh*-expressions are common, varied, and productive in Mesoamerican languages, to a degree that is often much higher than what is found in Spanish or English or any other language within the SAE group.

We have shown that all the main varieties of [−H]Rs are productively attested in Mesoamerican languages and each variety has to be investigated on its own, because of the important differences in the meaning and the morphosyntax we discovered (e.g., the different subset of *wh*-words that introduce different [−H]Rs). Finally, we have also shown that semantics plays a crucial role in identifying and defining different kinds of [−H]Rs.

---

33 See Dayal (2013) for a proposal to handle this behavior of bare nominals without assuming that they trigger existential quantification and Chierchia (in preparation) for a critique of this approach.
34 “Headless relative clauses (like English *what I bought at the store*) are not relevant to this map” Dryer (2013a: §1). Later in the same section, ex. (9a) provides an example of [−]HRC from Awa Pit. Dryer (2013b: ex. (34d)) contains an example of a [−]HRC from Nadëb. I have not been able to find [−]HRCs mentioned anywhere else in the database.
References


Chierchia, G. (2020). “People are fed up; don’t mess with them.” Non quantificational arguments and polarity reversals. Manuscript. Harvard University.


89(4), 607–624.