

Puzz-Ling

a hands-on model of syntax to popularize linguistics as a science

Luca Ducceschi and Roberto Zamparelli

Draft

Abstract

Modern theoretical linguistics is arguably a science with a well-defined methodology, relatively simple and inexpensive data-collection strategies and clear objectives, yet the study of human language is not widely recognized by the general public as a scientific endeavor. Formal syntax and semantics, for instance, have not been popularized like physics, chemistry or other hard sciences, and have very rarely found a space in science museums. In this paper we discuss some aspects of this problem, focusing on syntax and its potential in education, and arguing that the lack of popularization derives from the need for different methods of dissemination. To overcome this problem, we present a completely novel approach to syntax education, based on a physical, hands-on multilingual model which we have designed and built. The model, called Puzz-ling, consists of a set of tiles that can be combined only to form grammatical sentences, in three languages (English, German and Italian). A simple set of rules and symbols and a string-reordering system based on constituent swapping derives the correct word order in all three languages. The model introduces in a simple and entertaining form some key concepts of (generative) syntax (e.g. tree structures, binary branching, movement), and displays in an intuitive way the structural difference among languages. The system can be turned into a game with the addition of points and turns, resulting in a hands-on, “edutainment” approach to language. The mechanics of the game are briefly described, and various trade-offs encountered in the design phase are discussed.

1 Linguistics and scientific communication

College students asked to name a modern science might pick physics, biology, chemistry or any other “hard” science, all disciplines whose results have high visibility in the technology that surrounds us. Few, if any, would pick linguistics, even though its subject matter, language, is ever present and a defining feature of *Homo sapiens*. Modern theoretical linguistics is a science with a well-defined methodology, relatively simple and inexpensive data-collection

strategies and clear objectives, yet the study of human language (often found in the company of literary studies in academic departments, and frequently confused with them) is not widely recognized by the general public as a scientific endeavor. A reflex of this situation is the fact that linguistics is rarely represented in science museums, or in the best case, it is merged with sociology or with the study of other aspects of human cognition, despite the fact that the methodologies of psychology and theoretical linguistics are often widely different. The social, historical and cognitive dimensions are, of course, all important aspects of the incredibly multifaceted phenomenon that is language, and all have research traditions that hark back to the 19th century. However, the current situation disregards a huge part of modern linguistic research—all the work done since the 1950s in formal syntax, by scholars such as Chomsky, Bresnan, and Jackendoff, and in formal semantics, by Montague, Partee, Kamp, Krifka and many others. None of this research belongs to the public's perception of science, and all remains virtually unknown to most students, until they stumble upon introductory courses in linguistics at university level.

This situation stands in ironic contrast with the fact that language and grammar teaching figures prominently in the programs of most modern schooling traditions. In the European educational systems with which we are familiar (especially in our country Italy, as well as in Germany and Spain), the grammar taught in school is mostly prescriptive: the students learn a classification of construction types (say, the names of thirty-some types of complement clauses), without any attempt at explaining why these constructions should be as they are and not otherwise. In other words, students are asked to learn a theory which has no explanatory adequacy, makes no attempt at teaching critical reasoning on language facts, and typically does not try to do a systematic comparison of linguistic facts across different languages. The opinions on the effectiveness of this kind of grammar teaching on writing abilities and general education vary from country to country. In the US, the meta-analysis in Graham and Perin (2007) reached very negative conclusions, but these might in fact apply to the methodology, rather than the content of grammatical teaching per se (see e.g. Navarre Cleary 2014).

Starting from these considerations, we think that there is ample room to improve the perception that the broader community has of the subject matter of linguistics and stress the problem-driven aspects of its methodology. But if this has to be done at school level, we have to resort to educational methods that go above and beyond traditional instruction.

2 Redesigning grammar teaching

One of the main problems with the role of traditional grammar in education is that it is, in most cases, top down: rather than reasoning about language data, the student has to apply pre-defined categories to more or less naturally occurring sentences. This has two drawbacks: it fails to make use of negative evidence, and it does not provide a systematic, structural, cross-linguistic comparison. The usefulness of negative evidence is exemplified by the minimal pairs in (1). Three expressions are grammatical, the one marked with “*”,

ungrammatical. Comparing them, the student should start to notice the difference between subjects and direct objects, two complements which dependency grammars (in Europe, the starting point for much school-level grammar teaching, see Hays, 1964; Tesnière, 1966) place at the same level. Similarly, textbook examples of the contraction of *want to* to *wanna*, in (2), may be used to make students reflect on the position that fronted interrogative elements occupy in the various ‘surprise question’ counterparts given on the right.

The classic analysis for the ill-formedness of (2b) is that a marker of the original position of the pronoun “who”, the “trace” *t*, located between *want* and *to*, blocks their contraction. See Boeckx (2000) for an overview of the issue. The fact that the students might not arrive at a precise formulation of the source of these effects is not important: language is used here as a training tool for abstract thinking, and tinkering. Indeed, as pointed out by Honda and O’Neil (1993), and in Hudson (2004), syntax provides the perfect lab to learn scientific methodology: its data can be extracted from the language competence of the students themselves, the experiments can be formulated in minutes and run in seconds, at zero costs, but the chain of logical reasoning behind the data analysis can be quite complex. Next, consider the issue of language comparison. In Trentino, Italy, students of certain high schools study Italian, Latin, Ancient Greek, English and German. With the possible exceptions of Latin and Greek, there is no institutional occasion in which the structures of these different languages are systematically compared. Without this comparison, it is easy to dismiss language differences as a matter of purely lexical differences. Indeed, if an L2 learner just looks at (3), she could conclude that word order in the three modern languages is the same, with Latin the odd man out. Participles, in (4), however, show that the German word order is quite different from that of Italian and English, closer in fact to Latin, a point which is strengthened by embedded clauses, in (5), while interrogatives show once again the similarity between the two Germanic languages.

- (3) Gianni **ama** Maria
John **loves** Mary
Johan **liebt** Maria
Publius Gaiam **amat**

(4) Gianni **ha mangiato** una mela
John **has eaten** an apple
Johan **hat** einen Apfel gegessen

- (5) ...che Gianni **ha mangiato** una mela
 ...that John **has eaten** an apple
 ...dass Johan einen Apfel **gegessen hat**
- (6) Gianni **ha** mangiato una mela?
Has John eaten an apple?
Hast Johan einen Apfel egessen?

As argued in Giusti and Oniga (2011), comparison of this sort may help to put current and dead languages in the same theoretical perspective, and train students to look at them with the same tools and questions in mind, gaining a better understanding of both. Nor is the approach limited to dead languages. The ever-increasing presence in the classroom of native speakers of non-Indo-European languages gives opportunities in two directions: testing hypotheses about language universals by widening the cross-language comparison, and boosting the social status and self-confidence of first-generation immigrants, once their language comes to be seen as a crucial source of data points in a vast cross-linguistic comparison. They learn that, within linguistics, any language has to be taken as seriously as the national language of the host country.

2.1 Hands-on models for grammar?

A serious objection which may be raised against these ideas is that using grammar phenomena and language comparison as a ‘gymnasium of the mind’ might be fine in theory, but in reality it is simply too difficult and abstract, both for the students and the teachers. Of course, teachers do not have the time to study the features of every language spoken in class. Of course, island conditions on Wh- movement are beyond what a junior high-school student should be expected to endure.

The problem, we argue, is not in the topics but in the methods. Physics, arguably a difficult and abstract science, has developed a number of strategies to make learning easier and more pleasurable. Since the foundation of San Francisco’s Exploratorium (<https://www.exploratorium.edu/>), many “hard sciences” have developed hands-on models (Nersessian 1991, Wells and Hestenes 1995, Ma and Nickerson 2006), which make abstract ideas more concrete and allow educators to make them the focus of educational games. Moreover, in the age of digital content, many of these models attract users *precisely* because they are *hands-on* and not just *heads-in*: there is a component of physical manipulation that helps “thinking with your hands” and fosters memory (see Marshall 2007, Schneider et al. 2011). Models like these do not exist for the combinatorial aspects of language. Of course, formal and computational linguistics have the ILO, the International Linguistics Olympiads, and there have been recent attempts at bringing linguistics to science museums (Wagner et al. 2015). But successful though they are, these models cannot easily be exported to school teaching the world over.

2.2 The Puzz-Ling hands-on model of multilingual syntax: overview

We set out to address this situation. Our goal was to create a physical model of syntax, which could eventually be used in schools or science museums to make linguistics as a science more accessible to a wider audience, stimulate a playful, problem-solving approach to the subject, demonstrate the existence of nonlinear structures behind the linear appearance of language and showcase the possibilities of a constituent-based approach to grammar, against the shallow dependency-based approach, derived from Tesnière (1966), which dominates our Italian schools¹. We chose syntax because it is more regular than morphology, presents ample cross-linguistic variation and can be backed-up by a massive amount of theoretical work. In accordance with the idea of “edutainment” (education plus entertainment) or “serious games”, our model can be seen as a sentence-level Scrabble®-like game, in which a user connects word-sized pieces to assemble tree -structures which can then be read off as linear sentences. The model is provisionally called Puzz-Ling.

2.3 Goals

Our goals in designing Puzz-Ling were simple:

- A. to allow the model to generate **a reasonable number** of different linguistic constructions;
- B. to give the users the possibility of creating sentences with **several different meanings**;
- C. to generate sentences which are **as syntactically correct** as possible;
- D. to have **as few pieces** as possible, and to make them work in sets of **different complexity**;
- E. to create sentences in **more than one language**, and **to be able to compare their different structures**;
- F. to have **as few combination rules** as possible, and to require **no linguistics background** beyond a basic understanding of parts of speech (Det, N, V_{finite}, V_{participle}, etc.)

In addition, we had some not negligible practical requirements:

- G. To make the model **physically possible** to assemble.
- H. To make the pieces **solid, cheap and easy to build**

¹For a recent description of the dependency-based approach common in the Italian school tradition, known as “grammatical valenziale” (valence-based grammar), see De Santis (2016). See also Parisi and Antinucci (1973) for a predecessor linked to generative semantics.

Some of these desiderata present obvious trade-offs: having several constructions (A) and several meanings (B) in various languages (E) leads to having many pieces, against (D), and more concretely, (G): a physical puzzle with as many pieces as there are (high-frequency) words is not practical, or cheap. The solution was to factor words in three parts: a syntactic category (the “frame”), a lexical content (the “head”) and a set of selectional restrictions piece (the “selector”) expressing a word’s selectional restrictions, namely the type of categories to which it can connect (see Figure. 3 below). Adding words to the model means adding heads, but frames and selectors can be reused by many words, provided they carry the same color (syntactic categories are color-coded). Moreover, the current prototype (whose mechanics are patent-pending) uses a sliding ribbon on the head, which allows the user to select, and thus make “active”, one of several words. The words may belong to different languages (also distinguished by color, on the ribbon), but must all have the same part of speech (POS) and the same selectional behavior.

Point (C) above---aiming for linguistic correctness---means that sentence-building should never end up with an ungrammatical string; this tends to run against point (F). The game Scrabble® is an example of a word model that does not enforce (C): the correctness of the sequences of letters a player assembles is enforced *a posteriori*, by the use of a dictionary. This is obviously impossible with sentences, which are not a pre-determined set, especially given (A) and (B). Moreover, if the model is multilingual, the users might find themselves assembling sentences in languages they do not know. The first solution to this problem was to enforce grammaticality at the local level: like in a handmade jigsaw puzzle, pieces that, syntactically, should not form constituents simply do not lock. We adopted this solution at all levels: each frame, head and selector has holes and latches; in words or word parts that should not go together, these elements do not connect. We call these “hardware constraints”. Yet, local constraints are not enough: language has many non-local dependencies, from agreement to Wh-questions; for these, we had to adopt a system of simple symbols and rules, typical of board games (“software constraints”, i.e. structural rules not implemented in terms of the physical shape of pieces), which trigger and constrain the rearrangement of the pieces in the tree structure.

In keeping with desideratum (G) and to reduce the number of pieces, we had to limit grammatical choices. Thus, we elected not to use the first/second person, plural forms, unaccusative verbs with “be” auxiliary, or positions for topicalization or focus. Some of these limits could have been handled by implementing a mechanical system of feature percolation a-lá GPSG (Generalized Phrase Structure Grammar), but such a choice would have gone against desideratum (H).

2.4 Handling long-distance dependencies

The requirement of linguistic correctness meant that we had to create tree structures and instruct the users to linearize them. We implemented a system equivalent to a lexicalized set of phrase structure rules in a generative grammar, loosely based on Chomsky’s Principles and Parameters (PP) framework (Chomsky and Lasnik 1993). The main departure was in

the treatment of head and phrasal movement, which we needed for questions,, as in (6), relative clauses, and German verb-second phenomena. Rather than having a piece move to an empty position, leaving a trace, we had potential landing positions introduce elements bearing a (single or double) *down-arrow*. A software constraint then states that elements with double down arrows must be *swapped* (*can* be swapped, if the up/down arrow is single, as in *fuoco* ‘fire’ in Figure 3b) with a head bearing an up-arrow, which may also pied-pipe (i.e: carry along) the tree fragments it selects, giving the effect of phrasal movement. The swap is successful if both parts fit into their respective landing positions, and if any symbol on the active word of the moved piece matches a symbol of the same type in the landing position (this insures, for instance, that only Wh-elements end up in the Spec of interrogative CPs). Figure 1 illustrates schematically with Wh-interrogatives: after the software rules have applied, a null interrogative C head has (obligatorily) swapped with the auxiliary, and the trace, initially inserted in [Spec,CP] has (obligatorily) swapped with a Wh-DP.² In English and German yes-no questions, like those in 6(b-c), the empty C swaps with the auxiliary or the main verb, respectively, while in Italian (6a) the empty C has no down arrow, so no swap takes place.

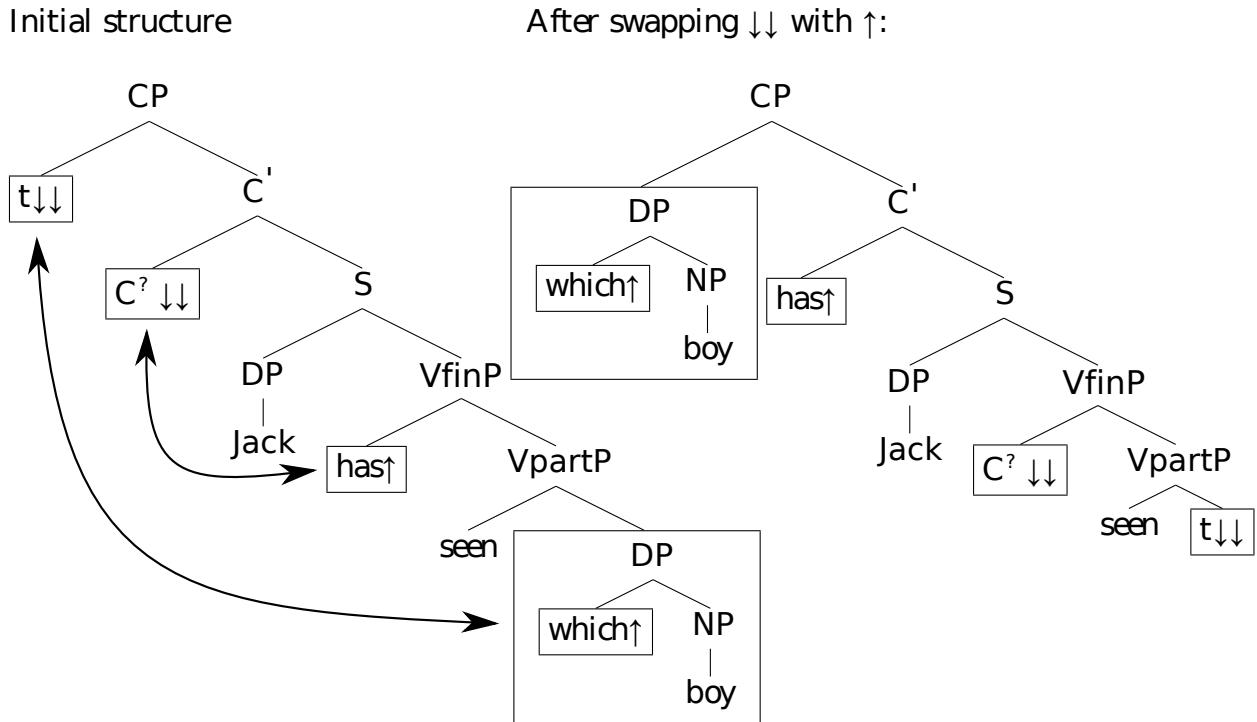
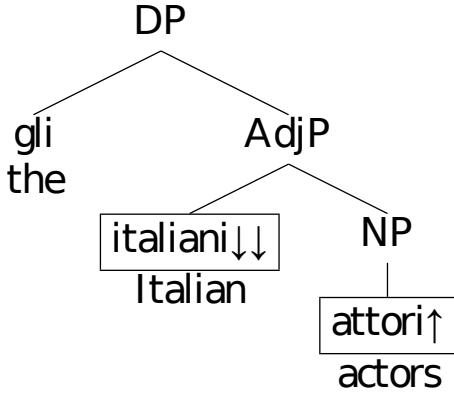


Figure. 1: A schema for wh-questions movement in English (head and phrase swapping). The labels for the various phrases ("P") stand for: "C" Complementizer, "D" Determiner,, "Vfin" Finite Verb, "VPart" Participle, "N" Noun.

²Fig. 1 might look like a classic D-structure/S-structure pair (Chomsky 1981), but the trees generated by the model and the swapping system are actually compatible with many derivational styles, including the bottom-up merge and move typical of “minimalist” approaches to syntax (Chomsky, 1995).

Figure 2 shows how the system can be used to invert word orders (the noun and adjective in Italian). This system allowed us to move elements to their correct positions without requiring users to generate empty specifier positions, which we found to be counterintuitive in practice and in contrast with the high-level constraint that for the whole structure to be well-formed every word with a protruding latch (see e.g. the right side of the selector in Figure 3 below) must find a connecting piece (there cannot be, in other terms, “unsaturated” words). This rule is partially relaxed only in the case of a third movement type, also guided by symbols, which plucks a head from a well-formed tree and places it *onto* a higher, target head. This is used for clitic movement (e.g. Italian “*Lui lo vede*” ‘He **it** sees’), and for combining certain bound morphemes, like English “-n’t”, with their bases.

Initial DP structure



After swapping Adj and N in Italian

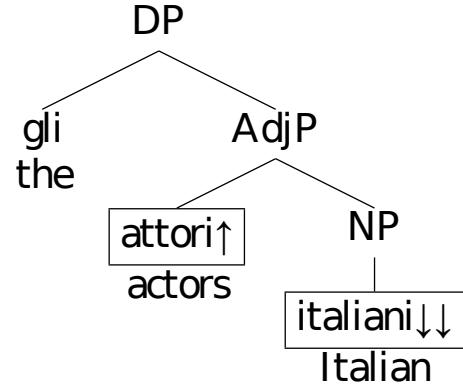


Figure 2: Head swapping in Determiner Phrases in Italian

Note that English nouns and adjectives will not have up/down arrows, so they won’t swap. More generally, cross-linguistic word-order variation is modeled in terms of presence or absence of arrow symbols, which trigger movement at the “software” level, altering the original structure. The main exception is the head-complement parameter, responsible for V(erb) O(bject) vs. OV orders, which will be described below.

2.5 Linguistic coverage

When all the pieces are used, the empirical coverage of the current version of the model includes several core syntactic phenomena, such as:

- **Determiner phrase (DP) structures** (Abney, 1987), with proper names, pronouns and determiners occupying the highest projection, D. We distinguish between nominative and accusative DPs.
- **Subject and object Wh-interrogatives** (with Subject-Aux inversion only in the latter, in English)

- **Verbs** selecting direct objects (“...drives **a car**”), participles (“...has **driven**...”), infinitives (“...must **drive**...”), gerunds (“...is **driving**...”), declarative CPs (“...thinks that **Sue left**”) and interrogative CPs (“...wonders **who left**”).
- **The complementizer layer**, with different frames for root clauses, declaratives and interrogative subordinate clauses.
- **Negation** and **NPI licensing**, plus some negation-driven object scrambling in German.
- Some **free relatives** (“what she bought scared her”)
- Some **passive** constructions.
- **SOV order** for German subordinate clauses
- The **verb-second position** for German root clauses (with subject or object in the initial position).
- Clitic **head movement** in Italian, and contracted negation in English (“-n’t”).

2.6 Mechanics

The Puzz-Ling model consists of a set of tiles that connect together to form words and fully grammatical sentences in three languages, English, Italian and German. In this section we consider some technical details of its physical implementation. The following videos show how some sample sentences are assembled; see the text below for a full discussion.

- Basic word assembly: <https://goo.gl/83J58w>
- Question formation: <https://goo.gl/sHcZPd>
- Some German sentences: <https://goo.gl/76Yh6Y>

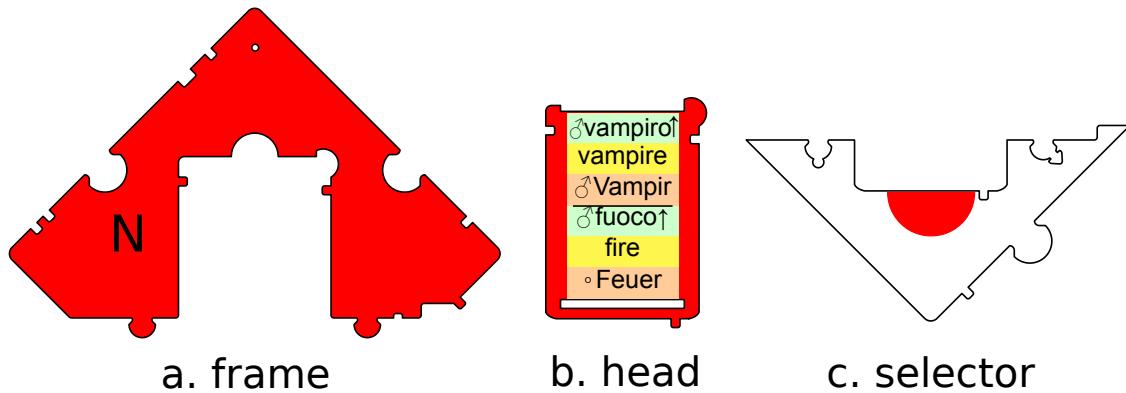


Figure 3: A word tile exploded in its three main components

As mentioned above, there are three main types of pieces in the model (see Figure 3), which result in a word when combined together:

1. **frames** (a), represent a word's syntactic category plus some grammatical features (e.g. Noun, Finite Verb, Gerundive Verb, Interrogative Complementizer, etc.);
2. **lexical heads** (b), rectangular pieces which fit inside frames and contain a sliding ribbon, which allows the user to select one lexical item out of several in the same category, in different languages;
3. **selectors** (c) express the selectional properties of the lexical heads (i.e: their compatibility with other frames; note that selectors can select other frames, not the heads inside them). The selector in Figure 3(c) has a latch protruding on the right, and will connect to another frame on that side. Selectors for heads that take no complements have no protruding latches.

These three components fit together as in Figure 4(a), forming a whole word. The available categories are nouns, verbs (finite, infinitives, gerunds, participles and passive participles), adjectives, determiners, negation, complementizers and prepositions. We do not have adverbials, which cause too many problems in SOV cases.

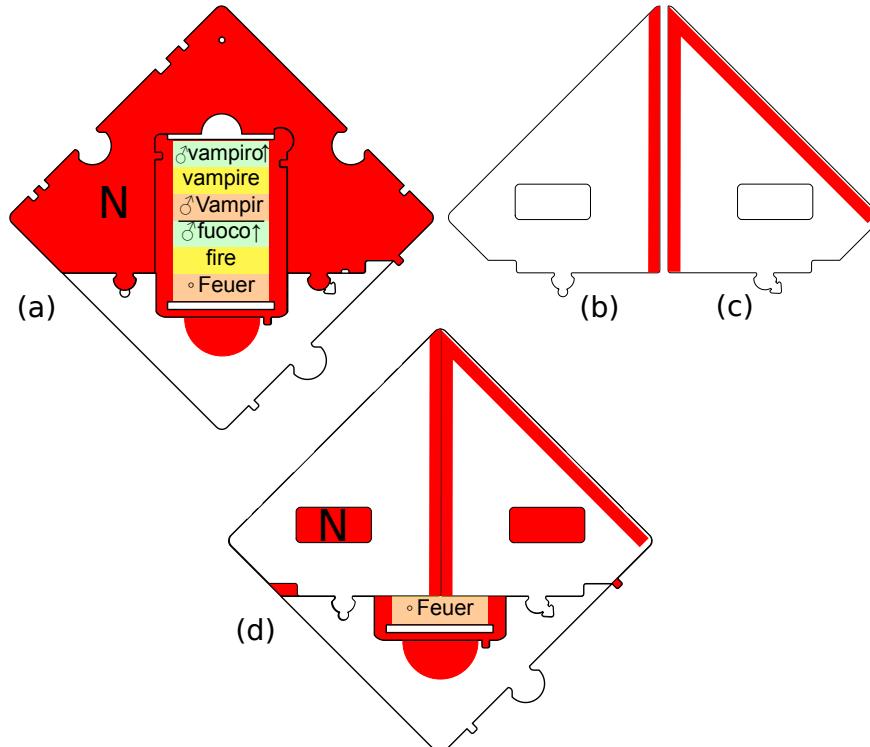


Figure. 4: a complete word (a), the two cover pieces (b-c) and the resulting fully assembled piece (d). The active word is the lowest (“*Feuer*”)

After the frame, the head, and the selector have been assembled into a complete word (Figure 4a), two pieces called “covers” (Figures 4b and 4c) are placed upon it (Figure 4d). The covers serve two functions:

- they hide all the items on the sliding ribbon except the lowest one, which becomes the *active* lexical item (“*Feuer*” in Figure 4d),
- the lines on the covers draw the left or right branches of the final syntactic tree, as in Figure 4(d). The choice of covers of type 4(b) or 4(c) is fully determined by the nature of the selector.

The model contains other types of pieces:

- a trace [*t*], used for phrasal movement (see Section 2.4);
- a piece used to connect nominal relative clauses as adjuncts;
- a “spacer”, which extends the length of the branches to avoid overlapping pieces (see e.g. Figure 7).

2.7 English, Italian and German

The set of images below (real photos of the prototype) exemplifies how the model works in practice. Starting from the initial English sentence *the Italian boy must buy a game*, in Figure 5, the users can slide the word ribbons and reassemble the same set of pieces to obtain two additional languages. In these examples the structures have the cover pieces on, which result in a clearly visible syntactic tree. For pronunciation, the user is instructed to start from above the root and follow the branches counterclockwise, spelling the words as they are encountered, unless they are set between square brackets (e.g. the Italian “[*pro*]” subject or the trace piece). This gives the linear order (see Figure 9 below).

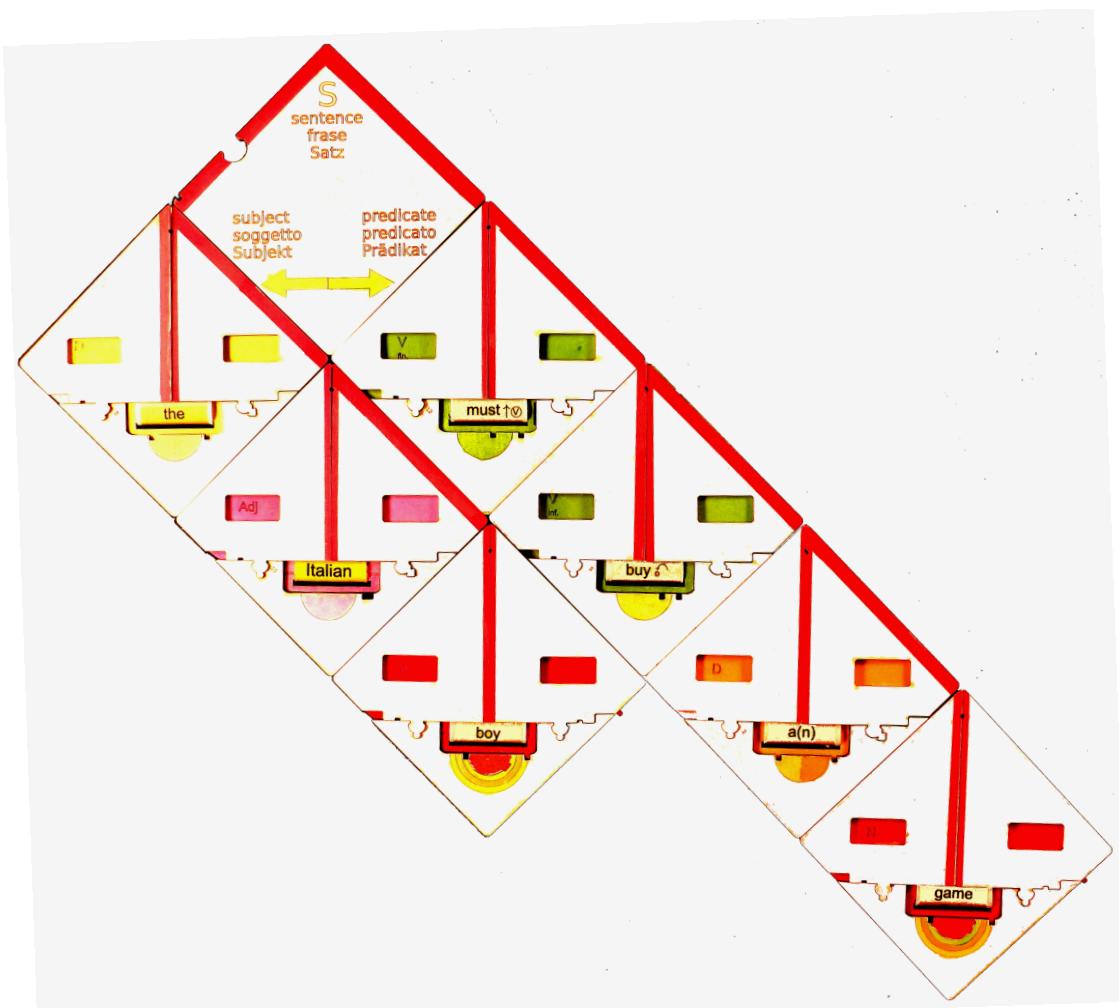


Figure 5: The English sentence: "*The Italian boy must buy a game*".

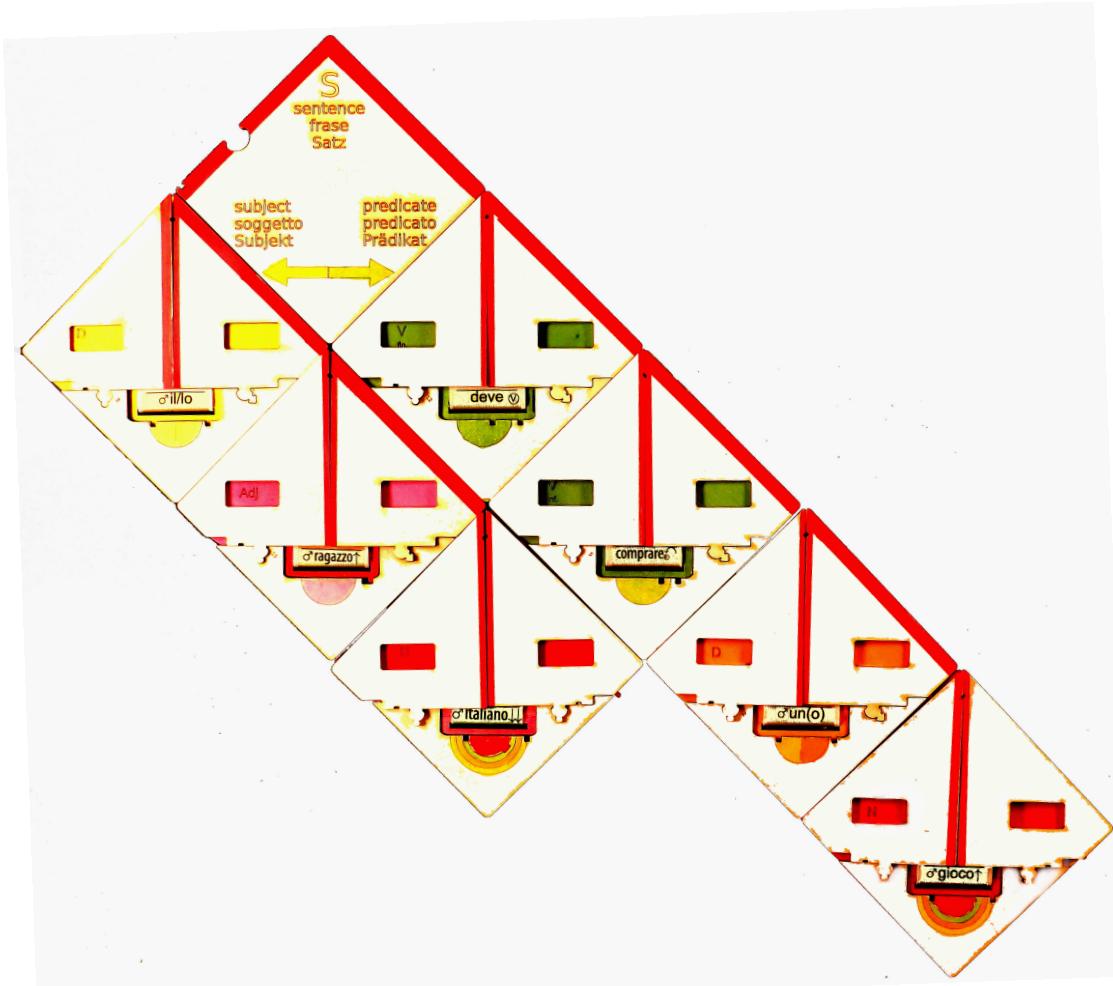


Figure 6: The Italian equivalent of the sentence in Figure 5: "*il ragazzo italiano deve comprare un gioco*"

The only structural difference between Figure 5 and Figure 6 is that adjective and noun have exchanged positions in Figure 6, as explained above. The differences between English and Italian become more evident in the case of negative or interrogative sentences, but even in simple sentences like these the user can make some valuable cross-linguistic observations.

The case of German is more complex, due to the combination of two factors: first, the Verb-second phenomenon, which, following den Besten (1983/89) is standardly analyzed as movement of the finite verb to C and of any phrase (argument or adjunct) to [Spec,CP]; second, the OV word order, which leads to the order Comp S O V_{non-fin} V_{fin} in dependent clauses. To generate OV structures, German verbs should be laid down as if in a mirror: any previously right-oriented complement thus becomes left-oriented. To obtain this effect, German verbs are written *upside down* in the word ribbon. The only way to align them correctly with the other words is to slide the ribbon to the other side of the head, then place the head, its frame, its selector and the covers on the face *opposite* to the one used by verbs in the other languages. This simple trick obtains OV order without any need for an

extra rule apart from the intuitive “don’t place words upside-down”. The situation with a dependent “...dass” clause is shown in Figure 7.

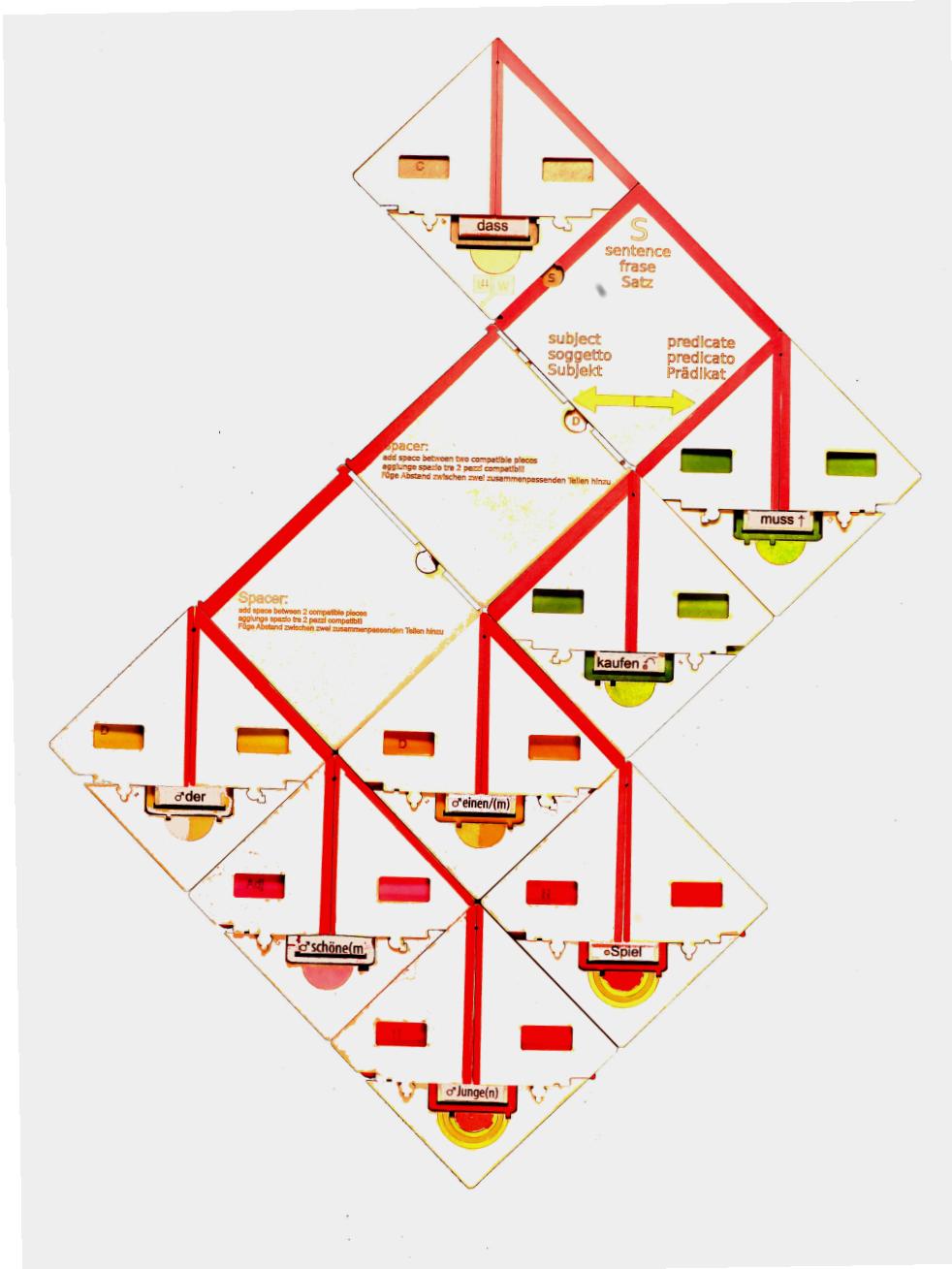


Figure 7: a subordinate clause in German, with order: S O V_{inf} Modal_{fin}: “*dass der schöne Junge einen Spiel kaufen muss*”.

Root clauses (Figure 8) differ in only one respect: the CP frame hosts a special null C head (restricted to German, as shown by its color) which: (a) swaps with any finite verb

in the clause, and (b) is only compatible with a selector that opens a trace-filled specifier. Unlike an interrogative specifier, here the trace can swap with any DP, subject or object, resulting in the German verb second root word order. With this, our model can cover all the constructions in the example sentences (3)-(6) above.

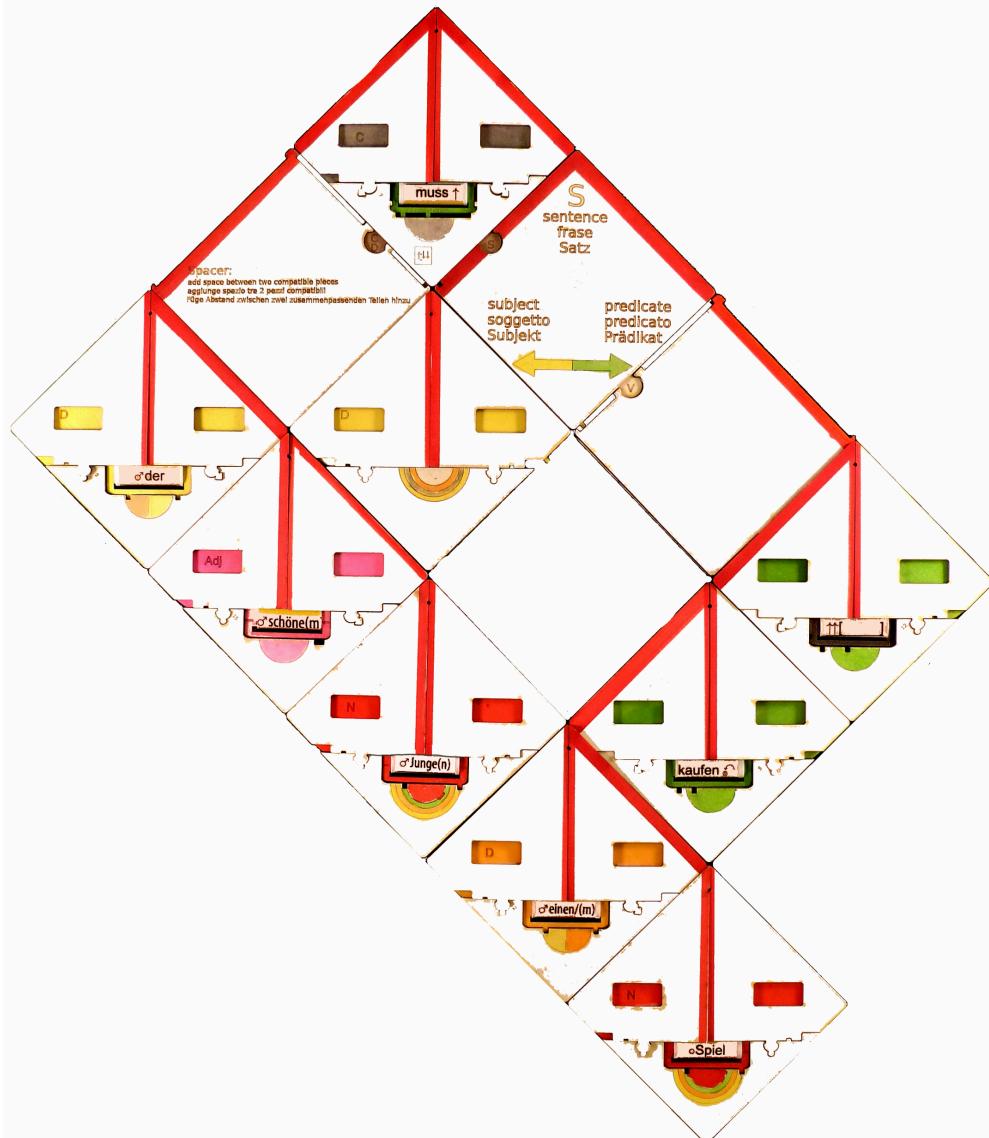


Figure 8: German Verb Second “*Der schöne Junge muss einen Spiel kaufen*” (the handsome boy must buy a game).

2.8 Limits of the models

While there have been other attempts at creating models of sentence composition, mostly in the domain of games,³ we believe Puzz-Ling to be the first physical model which (a) covers the range of constructions discussed in Section 2.5; (b) follows a well-established, coherent theory and (c) is multilingual without requiring different pieces. There are, of course, many aspects that remain well beyond its reach. We have mentioned the lack of plurals and sentential adjuncts; we add that the morphological realization of words, especially with cases different from nominative and accusative, are often not fully specified (in this case, the different options are given in brackets). While some aspects (e.g. island constraints) could be fixed via additional software rules, others (especially the fact that selectors select the frames, not the heads inside them) are clearly linguistically inadequate and force us to have a greater number of frames, against F. However, fixing these problems would require a much more complex mechanical system, against the desideratum (H). The current model was created with just plywood, glue, ribbon, markers, inkjet printing, laser cutting, and a lot of time.

In a way, comparing Puzz-Ling to a 'normal' syntactic theory misses the point. Puzz-Ling is not the implementation of a theory, but a "toy" model of syntax, whose main goal is to increase public interest in the complexity of language, the structural differences among languages and the problem-solving side of linguistics. It should turn on curiosity, not squelch it.

3 Beyond the model: turning syntax into an educational game

The mechanics of Puzz-Ling are undoubtedly complex, reflecting the complexity of the phenomena that it models. However, the game difficulty level---and the age group for which it is appropriate---can be easily tuned simply by classifying the pieces according to the complexity of the structures they produce, and reserving the more complex piece sets for more advanced players. For instance, by removing the complementizer layer and excluding German, the remaining system generates English and Italian declarative sentences, and becomes suitable even for the last years of primary school. To eliminate the problem of searching for the right combinations of frames, heads and selectors, which is time consuming, we have built a simplified version where these three parts are pre-joined together in a single piece (this means that there is no movement). For beginners, linearization remains a problem: we tackled it by explaining that the red lines in our model are fences, and that an ant placed

³See U.S.A. patents US 2520649 and US 4613309, both creating non-standard structures. Menza (2011) has an approach closer in spirit to ours, and applied to Latin, but (understandably, for Latin) he does not worry about word linearization.

on the topmost corner has to visit the whole tree, pronouncing the words as it finds them, without ever passing a fence, as shown in Figure 9.

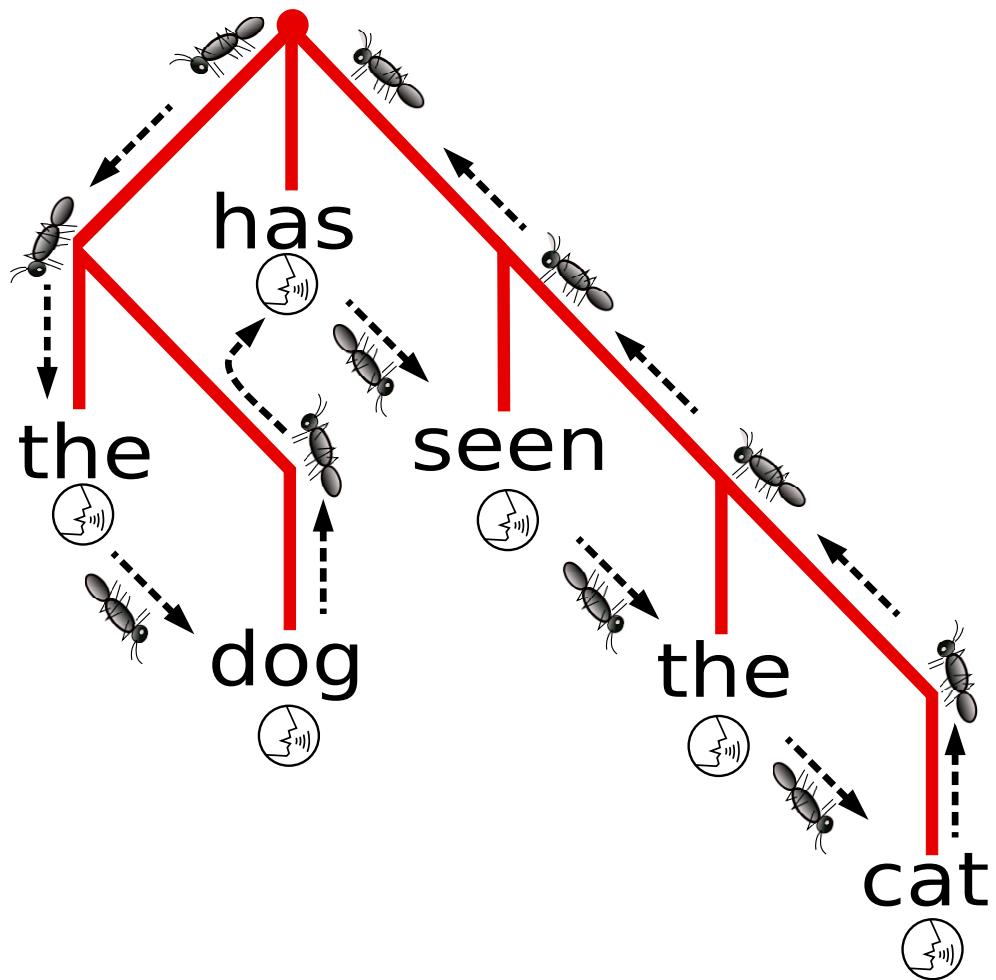


Figure 9: Linearizing "*il cane ha visto il gatto*" (the dog saw the cat): the talking ant walk.

Adding declarative Cs gives participants the possibility to build embedded sentences. We have found that the simple challenge “Build the longest sentence!” is usually quite effective to stimulate the players and familiarize them with the pieces. The next natural step is to allow German SOV dependent clauses. The mechanism that triggers a VO/OV inversion is probably the single most challenging step in the puzzle, and the players who solve it will probably remember the SOV order of German for quite some time. At this point, players have been exposed to the head swap mechanism with [Adj N] vs. [N Adj]; the next step is to introduce Wh-complementizers and generalize the swap mechanism to traces and Wh DPs. The final step is the head which trigger V2 and negation in German.

At present, we have showcased the model in various scientific events: two University of Trento “Researcher Night” open public events, the Drongo Language Festival 2016 (<http://www.drongofestival.nl/>), “Educa”, a festival of education targeted at primary and secondary schools (<http://www.educaonline.it>), plus various student-oriented events. The puzzle attracted quite a few people. Undoubtedly, this was initially due to its appearance, yet some remained a long time for the “build the longest sentence” challenge.

We tested it in an 8th grade class (4 sessions, 8 hours in total), collecting students’ impressions, and eventually asking them to draw tree diagrams (previously unknown to them) for simple sentences. We noticed a correlation between the ability to correctly place pieces in the puzzle and the ability to draw trees without looking at the puzzle. We were warmly asked by the teachers to repeat our hands-on lesson the following year. We also tested the simplified model on primary school students, under supervision. In this case, the interest in the game seemed to vary a lot from child to child, but sentences with outlandish meanings were a clear hit (remember that nothing in the model blocks *semantically* deviant arrangements).

These qualitative tests suggest the following considerations. First, delivering the model in steps, as sketched above, is very important, since it avoids overwhelming the player with hard-to-find pieces and difficult swaps at an early stage. Second, it would be useful to develop a competitive system based on points associated to both pieces and movements, in order to maintain engagement for those players who are not jigsaw puzzle aficionados. In class, we observed a drop in attention with the introduction of the more complex parts of the syntactic theory behind the model (movement, traces, V2), which could be avoided if the players are already experts when these pieces enter the game, and come to see them as an opportunity to gain quick points. The score system could be complemented by a card system which randomly assigns tasks (“+20 if you turn everything to German”, “Insert a negation or lose -20”), which would make players realize that, say, German has no progressive, and push shy players toward a more adventurous play style.

As with all hands-on models, the final and most pressing question is: to what extent can the model trigger a real understanding of its domain? Can it be used as a scientific discovery tool, or will it remain a mere game? And as with all such models, the answer is: it depends. Players who naturally engage with puzzle games will naturally develop questions. These strange rules--- do they really come from *language*? What would happen if I wanted to add *adverbs* to the game? For others, the presence of a game-master who has fully understood the mechanism and the phenomena it models will be essential. His or her task will be that of introducing the model in stages, as detailed above, but also providing sentences such as (3)-(6) and challenging the players to generate them, moving from language to the model, rather than in the opposite direction. If the game is ever going to be commercially produced, a future task we hope to carry out is to create a teacher textbook on the various ways of using Puzz-Ling. As usual, finding the right balance between entertainment and education will require practice and additional field testing, which we hope to carry out in the future.

4 Conclusions and further work

We have described Puzz-Ling, an educational, hands-on model for bringing multilingual syntax into schools and science museums, focusing on the generative side of language. Our hope is that the model can be adopted in education, and used to supplement current practices in linguistics teaching, showing another side of linguistic research. While we do not expect that practice with Puzz-Ling can directly enhance writing performance (we suspect that only hands-on writing can enhance writing), we do expect that it should lead to an increased awareness of language structures, and to a different approach to L2 learning and teaching. Indeed, we believe that Puzz-Ling has its greatest potential with teachers: mastering it to be able to teach with it may change teachers' views on how different languages work, and, we hope, lead to stimulating comparisons and interactions. For students, Puzz-Ling can be an entry point to understanding how languages differ, and which types of rules apply to them. Our greatest hope, maybe our dream, is that they could get to the point of wondering how Puzz-Ling could be improved: adding other constructions, solving its imperfections, converting it to other languages. As linguists, we owe them the thrill of playing with language.

5 Cited work

- Abney, S. P. (1987). The English noun phrase in its sentential aspect. PhD. thesis, MIT, Cambridge MA.
- den Besten, H. (1983/89) On the interaction of root transformations and lexical deleter rules. In: Werner Abrahams (ed.) *On the formal syntax of the Westgermania.*, 47-131. Amsterdam: John Benjamins.
- Chomsky, N. and Lasnik, H. (1993) Principles and Parameters Theory, in *Syntax: An International Handbook of Contemporary Research*, Berlin: de Gruyter.
- Graham, S., and Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, Vol 99(3), 445-476.
- Hays, D. 1964. Dependency theory: A formalism and some observations. *Language*, 40: 511-525. Reprinted in *Syntactic Theory 1*, Structuralist, edited by Fred W. Householder. Penguin, 1972.
- Hudson, R. (2004). Why education needs linguistics (and vice versa). *Journal of linguistics*, 40(1), 105-130.
- Klemmer, S.R., Hartmann, B. and Takayama, L. (2006) How bodies matter: five themes for interaction design. In Proc. of DIS '06, 140-149.
- Lillard, A. and Else-Quest, N. (2006) Evaluating Montessori Education. *Science*, 313(5795), 1893-1894.
- Ma, J and Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Comput. Surv.* 38(3), Article 7.

Menza (2001) Syntapuzzle. Un modello concreto della combinazione sintattica basata su proprietà lessicali e un materiale didattico di ispirazione montessoriana. Edizioni Youcanprint.

Marshall, P. (2007). Do tangible interfaces enhance learning? Proceedings of 1st international conference on Tangible and embedded interaction (TEI'07). 163-170.

McKee, C., Zimmer, E., Fountain, A., Huang, H.-Y., and Vento, M. (2015) Public Outreach in Linguistics: Engaging Broader Audiences. *Language and Linguistics Compass*, 9: 349–357. doi: 10.1111/lnc3.12153.

Nersessian, N. J. (1991). Conceptual change in science and in science education. In History, Philosophy, and Science Teaching, M. R. Matthews, Ed. OISE Press, Toronto, Canada, 133–148

Giusti, G. and Oniga, R. (2011). Why Formal Linguistics for the Teaching of Latin, in Formal Linguistics and the Teaching of Latin, Oniga, R., Iovino, R., Giusti, G., Eds., 1-22, Cambridge Scholars Publishing.

Schneider, Bertrand, et al. Benefits of a tangible interface for collaborative learning and interaction. *IEEE Transactions on Learning Technologies* 4.3 (2011): 222-232.

Tesnière, L. (1966). Éléments de syntaxe structurale, Klincksieck, Paris. 2nd Edition.

Wells, M, Hestenes, D (1995) A modeling method for high school physics instruction. *American Journal of Physics*. 63, 606