Human languages are extraordinarily complex adaptive systems. They feature intricate hierarchical sound structures, are able to express elaborate meanings and use sophisticated syntactic and semantic structures to relate sound to meaning. What are the cognitive mechanisms that speakers and listeners need to create and sustain such a remarkable system? What is the collective evolutionary dynamics that allows a language to self-organize, become more complex and adapt to changing challenges in expressive power? This paper focuses on grammar. It presents a basic cycle observed in the historical language record, whereby meanings move from lexical to syntactic and then to a morphological mode of expression before returning to a lexical mode, and discusses how we can discover and validate mechanisms that can cause these shifts using agent-based models.

This article is part of the themed issue ‘The major synthetic evolutionary transitions’.

1. Stages in language evolution

A human language is a remarkable, highly complex communication system. The capacity for language, the so-called language-ready brain [1], uniquely emerged in the hominin species, perhaps being in place as far back as half a million years ago [2]. Since then, languages have been born, and existing languages have kept changing, diversifying and dying. How can we develop a scientific understanding of the emergence and continuous cultural evolution of such a highly complex system? Analogous to a successful strategy in evolutionary biology [3], we could postulate different stages for the emergence of language in a population with language-ready brains, based on criteria related to the complexity of the meanings that can be conveyed and the complexity of the structures and linguistic forms available to express them.

To study how complexity at each stage arises and what is required to see transitions between stages, we could adopt the synthetic method, which is being used increasingly in many scientific fields, particularly biology [4], but also fields studying culturally evolving systems, such as sociology [5] or archaeology [6]. This method suggests that we should build operational models that generate analogous behaviours to those observed in the natural system we want to understand, similar to the way an aeroplane can be said to exhibit a similar capacity to fly as birds and hence informs us about what it takes to fly. In the case of language, the operational models take the form of a population of artificial agents which are initialized with a set of cognitive mechanisms and interaction patterns—but no language system—and after a significant series of interactions, usually called language games, we expect to see a communication system emerge that has similar properties as found in human languages, such as recursive syntax or rich conceptual structure [7–9]. The synthetic methodology is not only being applied using computer simulations [10], but also using physical robots (see figure 1 from [11]), so that the behaviour of the agents is embedded in reality and issues related to the perceptual grounding of language, and the relation of language to physical action can be addressed. In such cases, agent-based models resemble the artificial systems considered in synthetic biology [12], because they are embedded in the ‘real’ physical world.
The main advantages of the synthetic methodology are that (i) all internal states of the operational model can be tracked, for example, in contrast to humans, we can monitor the complete brain states of a robot as it is learning and processing language; (ii) experimental conditions can be varied in a controlled way, so that we can isolate the causal effect of a particular factor, e.g. increased population size, a new concept formation mechanism, increased communicative pressure; and (iii) a space of possible evolutionary linguistic pathways can be explored that have not necessarily occurred in nature, giving a theoretical tool for studying the space of possible languages.

This paper illustrates this methodology, focusing on the emergence of grammar. It is the final one of the three stages commonly recognized in language evolution research.

(i) From action to gesture. The first stage goes from purposeful actions, for example grasping, to symbolic communicative gestural signs, for example pointing, possibly accompanied by sounds to draw the attention of the listener. The gestures are not innate, but created and implicitly negotiated. This stage is reached by most children around the first year of life [13]. Many researchers have argued that gestural signs must have been the first stage in the origins of symbolic communication in our species as well [14], partly because closely related species, in particular chimpanzees and bonobos, also develop gestural signs among close kin [15]. Various agent-based models have tried to emulate this stage, mostly based on operationalizing ontogenetic ritualization [16–18].

(ii) From sounds to words. In the second stage, the sounds accompanying gestures, which were initially purely intended for grabbing the attention of the listener, become words, i.e. complex vocalizations associated with meanings [19]. This stage is reached by children in the first year of life with at first a slow acquisition rate, which then steadily increases so that around the age of 2, we typically observe a vocabulary spurt. Such a stage has also been postulated as the earliest stage in the origins of human languages [20], possibly already reached in earlier hominin species, such as Homo heidelbergensis [2]. Although some non-human primates can acquire a system of signs [21], these signs were always supplied by human experimenters as opposed to being self-generated, and systems do not propagate beyond close kin. Several agent-based models for stage II have been developed (see the sample in [22]). They typically take the form of language games in which agents from a population take turns being speaker and listener in order to refer to objects or actions in the shared reality, as illustrated in figure 1 discussed in Steels [11].

(iii) From single words to grammar. In the third stage, utterances use various syntactic devices, such as affixes, word stem changes, sequential ordering, intonation, stress patterns and hierarchical structure, in order to express additional meaning and extra information to avoid combinatorial complexity in parsing and semantic interpretation. This stage is reached in children by the end of the second year [23] and both lexicon and grammar grow rapidly until, at year five, the main grammatical systems of the language are in place, although many subtleties still need to be learned in subsequent years while the lexicon also expands further. Indeed, language learning goes on throughout life as languages are continuously changing. There is a wide consensus that humans are unique in their capacity to build up, learn and align grammatical language, even though many animal species exhibit some of the cognitive prerequisites such as recognizing and producing recursive syntax [24]. The rest of this paper discusses agent-based models for this third stage, which so far have been much less explored.

2. How grammar evolves in human languages

Historical linguists, like Wilhelm von Humboldt, already observed in the nineteenth century that the expression of meaning in a particular language is not static but cycles through different modes [25] (figure 2).
When looking at a particular language, we find that all four modes are used, although usually one of them is dominant. For example, English is predominantly an analytic language, i.e. primarily relying on syntax (mode 3), even though there are some meaning domains using a morphological approach, for example, past tense expression in irregular verbs (came/came, do/did) or expression of semantic roles (cases) in pronouns (he/him/his). Quechuan, a native American language, is predominantly a synthetic language, i.e. primarily relying on morphology (mode 4). Nouns have no less than 19 possible case suffixes, and seven possessive suffixes. Verbs have a variety of suffixes to indicate tense, aspect, mood and modality, and various markers to convey subtle aspects of meaning. For example, an action verb may have additional markers to express the nature of the action, how the action was executed, which type of instrument was used, which evidence was available, etc. [28]. All this information is expressed in English, using separate words organized in syntactic patterns.

The historical record shows that the language inventory tends to expand for each mode. We see growth in the lexicon, increase in morphological complexity [29], increase in the sophistication of syntactic patterns [30]. However, we also see weakening and erosion within a given mode that may lead to a shift between modes, a process that is commonly called grammaticalization [31–33].

A typical example comes from the domain of time, such as the expression of future [34]. We see languages where future is not expressed explicitly or very ambiguously, and it therefore has to be inferred (as is currently the case in Chinese). Then, a stage may develop where future is expressed syntactically with a verb phrase and an auxiliary (as in English ‘I will come’). Typically, a verb such as ‘want’ or ‘go to’, which indirectly suggests future, is recruited and then becomes grammaticalized to take on the role of a future auxiliary. Next, we may see the compaction of a phrasal pattern into a single word, as in French ‘Je partirai’ (I will leave), which comes from an earlier syntactic expression of future with the verb ‘habere’ (have) as in ‘partire haveo’ (literally, ‘to leave I have’). The words in this pattern were increasingly glued together and compacted to ‘parti-abeo’, then ‘partir-ayo’,
which finally ended up through further phonetic optimization as ‘partirai’ (French first person singular future of ‘partire’) [35].

Although mode shifts typically take place over long time periods (thousands of years), there can be sudden accelerations, often owing to catastrophic events causing punctuated equilibria [36]. For example, in the formation of creole languages, the grammatical structures of the source language (e.g. French) get stripped away almost entirely to yield a lexical language (mode 2 dominating) from which the grammaticalization processes start off again and quite rapidly (in one or two generations) lead to a novel fully functioning grammatical system [37].

Language evolution usually moves in the direction from lexical to syntactic and then morphological, but there are occasionally movements in the other direction [38]. For example, the form of a word may erode so strongly that the meaning is no longer clearly expressed at which point a re-invention takes place. For example, the Latin word for speaking ‘loqui’ went out of use in late colloquial Latin to be replaced by ‘fabulare’, which evolved into Spanish ‘hablar’, or ‘parabolare’, which evolved into French ‘parler’ [35]. The negation particle ‘ne’ in French, itself already a reduction from ‘non’, was felt as too weak and was reinforced syntactically with the particle ‘pas’ (literally step) as in ‘je ne veux pas’ (I do not want). Today, ‘ne’ has become optional and ‘pas’ has effectively become the negation particle, as in ‘je veux pas’ (colloquial contemporary French).

These uncontested facts suggest that we should not conceptualize the origins and evolution of language as an all-or-none phenomenon, perhaps owing to a single genetic mutation that gave rise to a single new operator (such as merge). Instead, it makes more sense to inquire about the many cognitive mechanisms, the invention, learning and alignment strategies, and the cultural evolutionary dynamics that have to be in place so that a population of individuals can sustain these linguistic cycles in the expression of meaning domains. The mechanisms include analogical inference, routinization of behaviour, optimization, analogical inference, hierarchical planning and plan recognition, concept formation, imitation, associative memory and no doubt many more. Whether these cognitive mechanisms are specific to a language, and therefore would have required neurobiological change, or not can only be discussed seriously if we have adequate operational models of what these mechanisms are.

3. Agent-based models of language evolution

Agent-based models are a good way to tease apart and investigate the many mechanisms and factors involved in explaining linguistic cycling, because we can explicate a certain factor or mechanism (e.g. a particular concept formation strategy or a particular mechanism for analogical inference) and study its effect on the emergence or change how a particular meaning domain (e.g. colour, space or time) gets expressed. Moreover, there are not only cognitive factors, but also external factors influencing the evolutionary dynamics; for example, strong language contact may lead to intensive borrowing and the subsequent collapse of phrasal or morphological patterns, significant population turnover may compromise cultural transmission, differences in frequency for meanings and forms may accelerate shifts to another mode or reorganization of the grammar [39]. All these factors can be incorporated in an agent-based model and their impact studied by systematically changing them, for example, by allowing slower or faster population turnover.

Much remains to be done, as this methodology is only now beginning to be applied on a sufficiently large scale, but there is already a body of significant case studies. Most importantly, research is converging on a core set of mechanisms and processes, so that the hugely complicated effort to set up an agent-based model becomes more doable. Our group has made concrete proposals for such a core and translated them into a computational workbench for doing evolutionary linguistics experiments that is freely downloadable from https://www.fcg-net.org/.

The common core in all our experiments includes the following components:

(i) A script-based interaction engine. It governs the turn-taking interaction between speakers and listeners.

(ii) A semantic representation formalism. Agents in all models use the same system for representing and manipulating meaning, based on a variant of second-order intentional logic. The representation includes objects in the domain of discourse denoted as symbols, e.g. o2, o3, etc., and n-ary predicates for the properties, relations or actions involving these objects, e.g. red(o2), next-to(o1,o2), etc. The semantic representations are second order, because a property or relation can itself be an object and the intension (the predicate itself) can also be an object. The details of this representation formalism are not important for the main points of this paper.

(iii) Representation of situation models. Situation models are couched in terms of this semantic representation formalism. A situation model contains all the objects and relations known about the shared context. It is private to each agent and not necessarily shared. In robotic experiments, the situation model is obtained from sensors and complex sensory processes that anchor objects in experienced reality and compute which predicates are true in the current context [40]. For example, a scene with a red ball that is inside a green box (which might occur in the experiments shown in figure 1) is represented with the following set of predications:

red(o1), green(o2), ball(o1), block(o2) and inside(o1,o2). (3.1)

(iv) Representation of utterance meaning. The meaning of utterances is obtained by the listener through parsing an utterance, and by the speaker through conceptualizing ‘what to say’ in order to achieve a particular communicative goal. The utterance meaning uses the same semantic representation formalism as situation models. However, expressions now have variables instead of constants. These variables are written as symbols with a question mark in front, such as ?x1, ?x2, etc. Semantic interpretation consists of matching utterance meaning against the situation model in order to find bindings for all the variables. For example, the utterance ‘ball red next-to block’ (assuming no grammar) gets translated by a lexical parsing process that looks up the word meanings and combines them into a set

ball(?x1), red(?x2), box(?x3) and inside(?x4,?x5). (3.2)
A possible binding for these variables, given the situation model in (3.1) is

\[ ?x1 = o1, \ ?x2 = o1, \ ?x3 = o2, \ ?x4 = o1 \quad \text{and} \quad ?x5 = o2. \]

Grammar primarily signals some of the co-reference relations between variables. For example, if we take the English utterance ‘(the) red ball inside (the) box’, then the listener knows, even before consulting the situation model, that \(?x1 = ?x2\), because the words ‘red’ and ‘ball’, which introduce the predicates ‘red’ and ‘ball’, are part of the same noun phrase. Moreover, he knows from the semantics of the prepositional noun phrase construction that

\[ ?x4 = ?x1 = ?x2 \quad \text{and} \quad ?x3 = ?x5. \]

(v) **A construction grammar engine.** Construction schemas are relevant for all modes, whether they capture lexical, morphological or syntactic ways of expressing meaning. A construction schema defines an association between meaning and form, for example the predicate ‘red(?x)’ and the word ‘red’. Or article, adjective and noun units are combined in a noun phrase that also establishes semantic relations between the meanings of constituents or additional predicates. In contrast to purely syntactic formalisms (such as minimalism), construction schemas always contribute meaning beyond the meaning of their constituents, such as co-reference relations between the meanings of constituents or additional predicates. Construction grammar typically organizes schemas into networks to support priming and inheritance from more abstract to more concrete constructions.

We have developed an operational version of construction grammar, called fluid construction grammar (FCG) [41]. Details of FCG are complex, but not crucial for the present discussion. The most important point is that FCG uses the same schema representation and the same processing mechanisms for lexical, syntactic and morphological constructions, so that agents can smoothly move between the three different modes of meaning expression described earlier. Another important characteristic of FCG is that the same construction schema can be used in parsing as well as production, so that the speaker can monitor his own speech by simulating how the listener might interpret the utterance he is producing, and the listener can simulate how he would express the meaning he was able to derive from the speaker’s utterance.

(vi) **Learning architecture.** Finally, all agents are equipped with a general architecture that supports insight learning [42,43]. There are two layers of processing: a routine layer where agents apply the constructions available at that point in their individual inventory, and a metalayer where agents apply diagnostics and repair strategies. A diagnostic strategy triggers when routine application of constructions is not possible, when the outcome after applying available constructions is incomplete or not interpretable with the current situation model, or when an opportunity for possible optimization is detected. A repair strategy attempts to deal with issues diagnosed by these diagnostics. For example, if a word is missing for expressing a fragment of utterance meaning, then the speaker may invent or recruit a new word; if a phrase or part of a phrase is recurring often, it may be compacted in a single word by changing the intonation structure and pauses between words; if a co-referential relation between two variables is not expressed, a grammatical construction might be introduced to convey this information to avoid semantic ambiguity in the future, etc. After an interaction, consolidation strategies come into action. They translate repairs into new constructions or variations on existing constructions, perform credit assignment and restructure the grammar.

(vii) **Cultural evolutionary dynamics.** All models we developed are based on an instantiation of evolutionary dynamics at the cultural, more precisely linguistic, level. Evolutionary dynamics requires that there is a population of units that multiply with inheritance, exhibit variation and undergo selection, effecting the distribution of traits. Here, the traits are strategies and constructions built with them, stored in the individual memories of the agents. They multiply through social learning as a part of language games. Variation is unavoidable owing to performance deviations, creative language use and imperfect learning. Selection takes place, because the agents prefer constructions that lead to higher communicative success, adequate expressive power and minimized cognitive effort, causing some strategies and constructions to propagate and become dominant in the language community. Language evolution never stops, because there is no optimal solution and no central authority, so that the population keeps navigating in the space of possible language variants to have a communication system adapted to their needs.

4. Case studies

From §3, it follows that setting up a specific agent-based model requires: a definition of a game script, an environment that will be the source of meanings, a population with possibly internal structure and dynamics and most importantly operational diagnostic, repair and consolidation strategies, so that we can see what their effect is on the emerging language. An experiment will typically focus on one aspect of language. These can be quite specific, for example, how can an inventory of colour terms and colour categories arise, how can a system expressing tense and aspect emerge, how does phonetic erosion lead to the collapse of a case system. It can also be more general, for example, how can a recursive phrase structure grammar emerge, what is the impact of analogy on streamlining an inflection system, what is the impact of population structure and renewal on the emergence and preservation of a lexicon. The remainder of this paper can only give a few concrete examples with details contained in the cited papers.

(a) Word formation

Agent-based models for word formation have to explain how implicit, inferred meanings can turn into explicit lexical expression. This happens in the experiment shown in figure 1 in which agents play naming games about visually
perceived objects in a shared situation using the diagnostic, repair and consolidation strategy shown in table 1.

Consolidation also needs to include credit assignment using the following strategy:

(i) In the case of a successful interaction, all new or modified constructions are stored and the scores of the constructions that were involved are increased, whereas competing constructions, i.e. constructions that could also potentially apply but lead to a dead end or to some other anomaly, are decreased, thus implementing the kind of lateral inhibition dynamics familiar from many neural network models, such as the Kohonen network.

(ii) In the case of a failed interaction, the scores of the implicated constructions are decreased, and new or modified constructions are not stored.

We see in figure 1 that these strategies indeed allow a population of agents to expand and share a lexicon. Words propagate but there is variation, because different agents may invent a new word when they do not have one. Progressively, some words win the competition based on the lateral inhibition dynamics and a shared lexicon with a minimal set of words for the meaning domain results. These strategies have already been applied for a variety of meaning domains, including experiments where both the categories and the lexicon evolve and get coordinated [44]. Experiments have also been conducted on word formation with multiword utterances (without grammar) using slightly more sophisticated diagnostic, repair and consolidation strategies [45].

(b) Syntax formation
Agent-based models for syntax formation have to explain how a set of words can get organized into hierarchical phrasal patterns. This happens in the experiment shown in figure 3 (from [43]) where agents play the syntax game [46], which is a variant of the naming game that requires the expression of n-ary relations, such as ‘inside-of’ or ‘moves-towards’, and hence ambiguity about the role of the arguments in the relation. Figure 3 is based on two strategies: one for building phrasal patterns by incorporating units, and another one for fitting existing words or phrases into an existing phrasal pattern through coercion [43].

The phrase-building strategy creates an extended phrasal pattern, which initially contains just a single word, by incorporating an additional constituent. For example, the pattern ‘blue block’ is extended, so that an extra property is expressed, as in ‘big blue block’. The new variant inherits most of its properties from the partially applicable construction (table 2).

The coercion strategy extends the lexical or phrasal categories of a unit, so that it fits a schema. For example, in the phrase ‘she WhatsApped me’, the noun ‘WhatsApp’ is coerced to behave as a verb, so that the clausal construction combining subject, verb and object, can apply (table 3).

Figure 3 shows the effect of these strategies when a population of five agents plays 800 syntax games. The experimental results show not only that agents converge on a shared phrase structure grammar (seen because alignment reaches almost 100%), but also that semantic and syntactic ambiguity decreases significantly to close to zero, implying that a ‘better’ communication system arises, meaning one with less cognitive effort and less risk for misunderstanding. Different agents may select different word orders to express the same co-reference relations, and so there is unavoidably competition. Agents may also differ on the grammatical categories of words. However, the lateral inhibition credit assignment strategy, the same as used for word formation, causes convergence of word ordering and categorical usage.
Agent-based models for the phonetic erosion of meaning have to explain how words or morphological markers of words can progressively simplify without compromising initially their use, until of course the form has completely disappeared. This is illustrated in the experiment shown in figure 4 taken from larger-scale experiments in the emergence of marker systems and agreement [47]. Simplifying, the speaker attempts to diminish the articulatory complexity of a form (with a certain probability) and the listener accepts this form-variant if it is close enough to an existing form and if its acceptance leads to a successful game, after which the listener consolidates the form-variant in a new construction as part of his own inventory. Results in figure 4 show this strategy at work with new variants popping up, and spreading in the population, without compromising communicative success, until they become dominant, after which a new variant comes up (table 4).

5. Conclusion
Agent-based modelling can play an important role in deconstructing the many factors that play a role in the emergence and evolution of language. It helps to tease these factors apart and study their causal impact on an evolving language evolution.
in repeatable objective experiments. Data from historical linguistics have abundantly demonstrated that there are different modes in which a particular meaning domain can get expressed (inferred, lexical, syntactic and morphological) and that we regularly see mode shifts creating linguistic cycles. For example, temporal information, such as future, may go from being unexpressed, to being expressed using words, auxiliaries in verbal patterns, and compact morphological expression in an inflectional system. A theory of language evolution needs to show by what cognitive mechanisms, external factors and evolutionary dynamics meanings get expressed in each of these modes and how shifts may occur.

This paper puts forward a common core with a script-based interaction engine, semantic representations for situation models and utterance meaning, an operational model of construction grammar processing, and a meta-level learning architecture enabling insight learning. It also puts forward a model of the cultural evolutionary dynamics and then points to a number of case studies that are using this core for achieving word formation, syntax formation and phonetic reduction. Clearly, language is an extraordinarily complex adaptive system. We therefore should not expect a single simple explanation how such a system can emerge and keep evolving, and many more experiments are needed on all aspects of the linguistic cycle.

**Competing interests.** I declare I have no competing interests.

**Funding.** The writing of this paper was partly supported by the European FP7 project Insight and by the Wissenschaftskolleg in Berlin.

**Acknowledgments.** The author is a fellow at the Institute for Advanced Studies (ICREA) and associated with the Institute for Evolutionary Biology (IBE) of the UPF and CSIC in Barcelona.

**References**