

An activation-based model of agreement errors in production and comprehension

Lars Konieczny (lars@cognition.uni-freiburg.de)

Center for Cognitive Science, IIG, Univ. Freiburg
Friedrichstr. 50, 79098 Freiburg

Sarah Schimke (sarah@cognition.uni-freiburg.de)

Center for Cognitive Science, IIG, Univ. Freiburg
Friedrichstr. 50, 79098 Freiburg

Barbara Hemforth (barbara.hemforth@lpl.univ-aix.fr)

Laboratoire Parole et Langage, UMR 6057, Univ. de Aix en Provence
29, av. Robert Schuman, 13621 Aix en Provence, France

Abstract

With this paper, we introduce the agreement production error (APE) model. APE is a model of comprehension and production performance that applies a theory of memory and cognition (ACT-R 5.0) to the task of linguistic processing embedded in a variety of psycholinguistic experimental paradigms.

With its roots in the ACT-R theory, agreement errors are modeled as a combination of symbolic processing and chunk activation dynamics. Whether a plural or a singular verb is produced depends on the accessibility of the Subject's plural marking. The activation of plural-marking chunks decays, so that it might not be found when its retrieval is attempted at the verb, resulting in a general singular error (Hemforth and Konieczny, 2003). This effect is then modulated by task and construction specific variations.

Introduction

When people speak or write they occasionally produce verbs not agreeing in number with the subject. This happens particularly often when the singular subject is followed by a plural modifier in constructions like (1; quoted from Bock & Miller, 1991).

(1) The readiness of our conventional forces are at an all-time low.

The mechanism underlying this error is attributed to the marked plural feature percolating up the tree too far (Vigliocco & Nicol, 1998). This account is substantiated by the fact that no comparable singular/plural mismatch effect for constructions with marked plural heads has been established so far.

Very recently, Haskell and MacDonald (2002) proposed the principle of proximity as an alternative explanation. They showed that in disjunctions like (2), subjects have a strong preference to match the number marking on the verb with the more local noun. In addition to distributional evidence, this was taken to indicate that

the classical attraction error at least partially and at least in English is caused by number marking on a close interfering noun.

- (2) a. The hat or the gloves is/are red.
b. Is/are the hat or the gloves red?

In a series of five written production experiments, Hemforth and Konieczny (2003) tested the proposed mechanisms in German.

In this paper, we follow up on this work and present a new model, APE, that accounts for the written production data in the experiments reported. The paper is organized as follows. We will start by summarizing the two most important experiments from Hemforth and Konieczny (2003). After that, the model will be outlined. Since APE is based on ACT-R 5.0, a short introduction to this theory is provided beforehand. The paper ends with a general discussion and conclusion.

Error patterns in written production

The first experiments replicated the classical results on subject-modifier-verb constructions. Two factors were varied in the first experiment: The factor "Match": matching (1,4) or mismatching (2,3) number marking on head noun and local noun, and the factors "Number of the head noun": singular (1,2) or plural (3,4) head noun.

- (1) Die Farbe auf der Leinwand _____ trocken.
The color on the canvas _____ dry.
(2) Die Farbe auf den Leinwänden _____ trocken.
The color on the canvasses _____ dry.
(3) Die Farben auf der Leinwand _____ trocken.
The colors on the canvas _____ dry.
(4) Die Farben auf den Leinwänden _____ trocken.
The colors on the canvasses _____ dry..

Hemforth and Konieczny (2003) found a clear effect of the number of the head noun on the percentage of agreement errors. Neither the factor "Match" nor the

number x match interaction reached significance. However, whereas no difference in matching versus mismatching local nouns could be established for sentences with plural marked head nouns, planned comparisons showed an effect with singular marked head nouns. This result replicates the well-known modifier attraction effect (e.g. Vigliocco & Nicol, 1998, Bock & Miller, 1991).

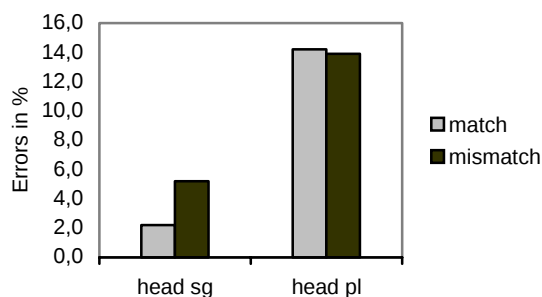


Figure 1: Agreement errors for NP-PP-V constructions

In line with earlier experiments on written production (e.g., Branigan et al., 1995; Hölscher & Hemforth, 2000), we found a considerably high number of agreement errors for plural marked head nouns, reflecting a general tendency to produce singular verbs. Nevertheless, the result for singular heads is compatible with both the feature percolation hypothesis and proximity. Hemforth and Konieczny (2003) therefore ran a series of experiments on Subject-Object-Verb (SOV) constructions. An object attraction effect for singular subjects would rule out feature percolation, because the object is not embedded within the subject.

Object attraction?

The experimental factors varied in three further experiments were “Match”: matching (9,12) or mismatching (10,11) number marking on Subject NP and local object NP, and “Number of Subject”: singular (9,10) or plural (11,12) Subject NP.

- (9) Ich habe gehört, dass der Mann die Frau besucht _____.
I have heard that the man(masc,nom) the woman visited _____.
(10) Ich habe gehört, dass der Mann die Frauen besucht _____.
I have heard that the man(masc,nom) the women visited _____.
(11) Ich habe gehört, dass die Frauen den Mann besucht _____.
I have heard that the women the man (masc, acc) visited _____.

- (12) Ich habe gehört, dass die Frauen die Männer besucht _____.
I have heard that the men the women visited _____.

In all three experiments, the number marking on the Subject had a strong effect on the number of agreement errors: more errors were produced following a plural Subject. However, less errors were produced when the local Object-NP was also plural marked.

The lack of an object attraction effect for singular subjects is consistent with the feature percolation hypothesis and contradicts proximity.

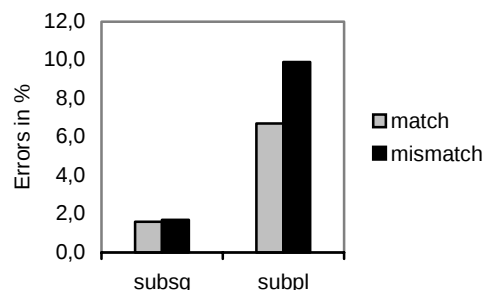


Figure 2: Agreement errors in SOV-constructions

The mismatch effect for plural subjects, however, is not predicted by feature percolation. Hemforth and Konieczny (2003) proposed *feature reactivation* as an explanation. According to that hypothesis, the plural feature of the subject (head) is subject to activation decay so that activation can be below the retrieval threshold when the verb must be produced. This mechanism would account for the general tendency to produce singular verbs. In SOV constructions, however, the subject plural feature can be reactivated by an object plural feature, because both subject and object are arguments of the verb.

APE: A hybrid model of agreement errors

The data so far suggest that both syntactic constraints and proximity affect agreement errors. For one, there are certain effects restricted to certain syntactic constructions (“feature percolation”), and second, there are effects of locality and interference best dubbed in terms of decay and reactivation. The Agreement Production Error (APE) model is built atop the ACT-R architecture, which provides us with mechanisms for *i.* declarative chunk activation and decay, embedded in a *ii.* symbolic processing architecture with *iii.* cost-dependent rule selection, and *iv.* task-specific modelling.

An informal introduction to ACT-R 5

ACT-R (Anderson and Lebiere, 1998; Anderson et al., submitted) is both a theory of cognition and a modelling framework, where scientists can build their specific models using well defined and empirically justified concepts that serve as the model's building blocks.

ACT-R distinguishes declarative from procedural knowledge. Both employ advanced sub-symbolic mechanisms.

Chunks are the elements of *declarative memory*. They bundle information in a collection of attribute value pairs, by which chunks get linked to other chunks to form networks of declarative knowledge. When chunks are created, they start out with a certain base level activation that decays over time, following the power law of forgetting. If their activation falls below a certain threshold, their chance for being retrieved by a production approaches zero.

When a chunk is retrieved from declarative memory, its base level activation is permanently pushed a bit upwards. The more often it is used, the higher it will be activated (power law of learning). In addition to base level activation, a chunk can temporarily receive additional activation spreading from associated chunks stored in the goal of the current production rule. The rationale is that chunks relevant for the problem stated in the current goal should benefit from being within the current focus of attention.

Production rules interact with declarative memory via retrieval of chunks. A retrieval request may succeed or fail, depending on whether or not a matching chunk exists, and on that chunk's activation. Among several matching chunks, the one that is activated highest has the highest probability of being retrieved. A retrieved chunk is stored in the retrieval buffer, where the next production can use it. Given a certain constellation of chunks in the available modality specific buffers (goal, retrieval, visual, audio, etc.), multiple production rules might apply to this state, but only one will be picked for firing. In ACT-R, conflict resolution in the case of multiple potential production instantiations is determined by the *utility* value of each rule in the conflict set. The *utility* is in turn a function of the past history of success and the cost associated with solving the problem by picking this rule.

If there are two productions in the conflict set, one of which has a high probability of being successful but which takes a while for getting there, and the other is quick but has only a mediocre success history, the choice will depend on the amount of time the user devotes to the problem. If the focus is on accuracy, plenty of time can be spent solving the problem so that the more successful rule will be picked. If the focus is on speed, the faster rule will be picked even though it's not unlikely that it will fail.

Basic modelling decisions

While ACT-R provides only limited means of representing and processing declarative and procedural knowledge, there are still many alternative ways in which knowledge can be modelled. We settled on the following modelling decisions before we actually started:

- Restricted use of direct storage. We do not store linguistic elements in slots of goals unless it is required for declarative reasons. Newly created chunks are released into declarative memory and retrieved when needed. Binding a chunk to a slot in the goal for procedural reasons is too strong a computational assumption, since stored chunks are excluded from activation decay.
- As a consequence of this, syntactic nodes have to be retrieved from memory in order to become integrated with other nodes. For instance, when the verb is processed, all its complements and adjuncts must be retrieved from memory to form an integrated interpretation.
- The cost of integrating a word is hence a function of the accessibility of its dependents in memory. Integration will be the easier, the more locally its dependents have been processed beforehand (cf. Gibson, 1998)
- The restricted use of direct storage has also consequences for the agreement mechanism, because NPs have to be retrieved all the time to get attached to modifiers or for incremental interpretation. The continuous retrieval of NPs influences their activation, so that some will be more accessible than others when the plural feature is to be assigned.

The sentence processing mechanism

The current model version performs the completion task as used in Hemforth and Konieczny (2003). First, a series of noun phrases are processed before the model produces a verb.

Processing starts with a goal that represents the current processing state during the assembly of the sentence elements. Each word read from the screen triggers the retrieval of a lexical element and is integrated into the currently processed phrase marker. The first NP is marked as the subject of the clause. When a modifier is processed, its host will be retrieved for attachment.

At each top level element of the sentence, a propositional interpretation is sought that matches the concepts associated with the processed phrases to a long term proposition (cf. Budiu & Anderson, 2000). If such an interpretation can be found, the concepts are hooked to that proposition.

In verb-final constructions, multiple arguments might precede the verb. In the absence of thematic

information of the verb, APE anticipates the verb by retrieving a proposition in the background knowledge that integrates the arguments processed so far. The more arguments have been read, the more likely will the interpretation match the right one when the verb is read, so that actually integrating the verb will become easier (Konieczny and Döring, 2003). In this view, anticipation is at the conceptual rather than the syntactic level.

Modelling agreement

Representation of number. Singular is the default number of nouns, whereas plural has to be assigned explicitly. A noun phrase is hence considered singular unless it is marked plural.

The plural feature is modelled as a chunk that links the plural attribute to a noun phrase, rather than as a slot in an NP chunk. As a chunk, a plural feature is subject to ACT-R's activation dynamics.

Producing number. Verb production is modelled by distinct production rules for singular and plural forms. The plural rule attempts to retrieve the plural feature of a noun phrase marked as subject of the sentence. If it succeeds, the plural form is generated from the base form. If it fails, the singular rule produces the singular base form.

The plural rule has higher utility, due to its better accuracy in actually producing the right verb when the subject is plural. The singular rule is less specific and therefore error prone, but less cost intensive.

Where syntax matters. The German production data suggest that plural attraction is construction specific: While there is a robust modifier attraction effect for singular subjects in SMV constructions, there is none in SOV. This result has been predicted by the *feature percolation hypothesis*.

Feature percolation is an encoding error by nature, in which the plural feature is erroneously assigned to the head noun instead of the embedded noun.

ACT-R, however, lacks a direct mechanism for encoding errors (i.e. the creation and release of false chunks). The only place where errors can occur in ACT-R is during retrieval of chunks.

Modelling "Feature percolation". In APE, the construction specific encoding error is modelled as a retrieval error during the search for an NP that the plural feature is to be assigned to. That is, during plural assignment, the newly created plural feature requires a root NP that has to be retrieved from memory. At the modifier in a complex-NP construction, the head NP at that point is highly activated due to the fact that it had been retrieved for modifier attachment shortly before plural assignment. Since both the head NP and the modifier NP are about equally strongly activated, there is a certain chance that the head NP, not the modifier NP is retrieved for plural marking. If that happens, the

subject has inherited the plural marking from the embedded NP.

In SOV constructions, no attachment takes place between the Subject and the Object-NP. Therefore the Subject NP is not going to be retrieved before the plural assignment of the Object. Retrieving the root NP for the object plural feature is hence less error prone.

On the other hand, the subject NP will be retrieved *after* the object has been assigned to the new plural feature, because subject and object are both needed for incremental interpretation, i.e. the anticipation of a matching relationship between both. The interpretation depends upon the number of the entities to be integrated, as the examples () and () illustrate.

(8) Gestern haben die Professoren den Studenten ...

Yesterday have the professors_{nom} the student_{acc}
"Yesterday, the professors have ... the student."

(9) Gestern hat der Professor die Studenten ...

Yesterday has the professor_{nom} the students_{acc} ...
"Yesterday, the professor has ... the students."

Things that multiple professors are likely to do to a single student (examined, rejected, etc.) can be different from things that a single professor is likely to do to multiple students (taught, etc.). Number is hence an important feature at the conceptual level and useful for interpretation anticipation.

Differential plural re-activation in SOV constructions.

During the process of incremental interpretation of each new verb-complement, each previous concept participating in the proposition, and, importantly, its plural feature - if it exists - will be retrieved and hence its activation will be pushed a bit. That is, when the object in SOV constructions is processed, the plural marking of the subject will be re-activated. This will only happen, however, if it can be successfully retrieved when the object is interpreted. There is a slight chance that the plural feature cannot be found here because its activation has already decayed too strongly. This chance, however, will be lower, if the plural feature receives additional activation from the goal, which is the case if the object is marked for plural. The additional amount of source level activation results in a higher retrieval probability for the subject plural marking if the object is plural. Therefore, the subject plural will be reactivated more often, so that it can be retrieved better and more often when the verb is produced.

Implications

Whether a plural or a singular verb is produced depends on the accessibility of the Subject's plural marking. The activation of plural-marking chunks decays, so that it might not be found when its retrieval is attempted at the verb, resulting in a general singular error (Hemforth and Konieczny, 2003). This effect is then modulated by task and construction specific variations. The model will

come in variants for different experimental paradigms, which are nevertheless based on the same core for verb-production. The first model variant presented here performs the completion task for written production and is hence a combined sentence processing and production model. It first reads two NPs, embedded in a variety of constructions, and then produces a verb. Modifier attraction errors (cf. Bock & Miller, 1991; Vigliocco & Nicol, 1998) are due to encoding errors during plural marking (feature percolation). The plural feature of the modifier-NP sometimes gets wrongly assigned to the Subject-NP. This effect is due to the necessity of reactivating the Subject-NP in processing the modifier-NP and is therefore restricted to modifier-NPs (Hemforth & Konieczny, 2003). In SOV constructions on the other hand, at each new NP, all verb arguments get reactivated to allow incremental interpretation and verb-anticipation (Konieczny and Döring, 2003). During this process the plural feature of an Object-NP can reactivate a plural feature of the Subject, reducing the singular error in S-O-V constructions (Hemforth & Konieczny, 2003).

The model is currently extended to other types of tasks to be able to account for task-specific-differences.

Aural repetition and completion. In this task, the participants are first presented with a preamble and then have to repeat and complete it with a verb in order to form a full sentence. This type of task has been used in the majority of experiments on agreement errors (e.g. Bock and Miller, 1991, Franck, Lassi, Frauenfelder and Rizzi, 2003, Bock and Cutting, 1992, Hartsuiker, Antón-Méndez and van Zee, 2001, Vigliocco and Nicol, 1998) and is generally claimed to test “pure” production, because subjects actually utter the whole sentence and not just the verb as in a completion task. Nevertheless, there is an inevitable comprehension component in this task as well, as the preamble has to be presented to the participants in some form. Instead of accounting for only the production part, we model the entire task, including reading and memorizing the preamble, repeating and completing it. Processing the preamble entails forming declarative representations for phonological and syntactic/conceptual information as in the completion task. To repeat the preamble, the model will have to retrieve either its phonological or syntactic representation. After repetition of the preamble, the verb should then be produced in the same way as in the present model. The difference between this task and the written completion task is in the higher activation of all elements of the preamble, as it is not only read once, but than (partly) retrieved and repeated. This should have no influence on the encoding error, but should make the general singular error much less frequent, which is in fact what has been observed in the experiments above. Moreover, the model predicts a floor effect for length variations of the intervening

material in this paradigm, which is in fact what has been found (e.g. Bock and Miller, 1991, Bock and Cutting, 1992).

Time pressure. In the experiments reported by Hartsuiker, Antón-Méndez and van Zee, 2001, participants had to perform the production task under time pressure. The model predicts the observed effects: Under time pressure, productions that consume too much time are more likely to be ignored in favor of faster but less accurate productions. In particular, the plural-feature will sometimes be left underspecified for its root. Under-specified plural-features will then erroneously be retrieved at the moment of the production of the verb. This would happen even with plural-features of intervening object noun phrases, hence accounting for the object-attraction-effect obtained by Hartsuiker et al. Moreover, time pressure should increase the general singular error due to a change in the utilities of the productions which determine the number of the verb: the more accurate, but also more time consuming plural rule will have a lower utility than in other paradigms. This will lead to the singular rule firing more often, which produces the singular base form without checking for the plural feature.

Dual task paradigms. Finally, the model can also be extended to dual-task-paradigms, as applied by Fayol, Largy and Lemaire, 1994, and Hemforth, Konieczny and Schimke 2003, in which participants have to perform a second working-memory consuming task in addition to the sentence completion or production task. The cognitive load created by this second task will lead to less attention, i.e. source activation, being devoted to relevant chunks during the processing of the preamble. This will make both the encoding error and the singular error more likely, because they are both due to retrieval errors that become more likely if decay is stronger or starts from a lower level. As predicted, a higher overall number of errors was found in the experiments cited above.

The influence of a specific experimental paradigm may further interact with properties of the language which is investigated. Such an interaction may eventually explain a cross-linguistic difference which has been observed in SOV-constructions: In contrast to the German results reported above, several studies conducted on French SOV-construction found a mismatch effect in the SP-condition (“object attraction”, see for example Fayol, Largy, Lemaire, 1994; Franck, Lassi, Frauenfelder, und Rizzi, 2003). In French, if there is an object in preverbal position, it always has to be a clitic pronoun. As these pronouns are very short, they have to be processed in less time than a full NP. Under such conditions, the model would predict the same effect as under time pressure: the plural-feature may be left underspecified for its root and

might then be retrieved when the verb has to be produced, leading to a plural-error.

Conclusion

We have introduced APE, an activation-based model of agreement errors in production. The model emphasizes the activation dynamics of the plural feature as the major source of variability in the data. Syntax effects are accounted for via operations that some constructions require that others do not. For instance, while modifiers have to be attached to their hosts, objects are not attached to subjects. On the other hand, objects, like other arguments, participate in anticipating the verb, while modifiers to NPs do not (to same extent, at least). These construction-specific operations interact with the activation dynamics of the plural feature in systematic ways that have been demonstrated to cover a wide variety of agreement phenomena discussed in the literature. The model predicts that time and feature re-use are crucial variables in production. Unlike purely linguistic production models of agreement errors, it can account for differences in task demands and non-linguistic factors of agreement performance.

Acknowledgments

We would like to thank Rob Hartsuiker, Michel Fayol, Julie Franck, Gerhard Strube and two anonymous reviewers for their helpful comments on earlier versions of this work.

References

- Anderson, J. R., Lebiere, C. (1998). The atomic components of thought. Lawrence Erlbaum Associates
- Anderson, J. R., Bothell, D., Byrne, M. D., & Lebiere, C. (submitted). An integrated theory of the mind. *Psychological Review*.
- Bock, K., & Miller, C.A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45-93.
- Bock, K. und Cutting, J.C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99-127.
- Branigan, H. (1995). *Language processing and the mental representation of syntactic structure*. Unpublished doctoral dissertation, Edinburgh University.
- Budiu, R. & Anderson, J. R. (2000). Integration of background knowledge in sentence processing: A unified theory of metaphor understanding, semantic illusions, and text memory. In *Proceedings of the Third International Conference on Cognitive Modeling*, pp. 50-57. Groningen, Netherlands: Universal Press.
- Fayol, M., Largy, P. und Lemaire, P. (1994). Cognitive Overload and Orthographic Errors : When Cognitive Overload Enhances Subject-Verb-Agreement errors. A Study in French Written Language. In: *The quarterly Journal of Experimental Psychology*, 47A (2), 437-464.
- Franck, J., Lassi, G., Frauenfelder, U. und Rizzi, L. (submitted). Agreement and movement: a syntactic analysis of attraction.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Hartsuiker, R., Antón-Méndez, I., & van Zee, M. (2001). Object-attraction in subject-verb agreement construction. *Journal of Memory and Language*, 45, 546-572.
- Haskell, T., & MacDonald, M. (2002). Proximity does matter. *Paper presented at the 15th annual CUNY conference on human sentence processing*, New York, March 2002.
- Hemforth, B., & Konieczny, L. (2003). Proximity in agreement errors. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, August 2003, Boston, MA.
- Hemforth, Konieczny, & Schimke (2003). Modifier attraction and object attraction. *Poster presented at the conference on architectures and Mechanisms of Language Processing (AMLaP)*, August 2003, Glasgow.
- Hölscher, C., & Hemforth, B. (2000). Subject-verb agreement in German. In B. Hemforth & L. Konieczny, *German sentence processing* (279-310). Dordrecht: Kluwer Academic publishers.
- Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads. Evidence from eye-tracking and SRNs. In: P. P. Slezak (ed.): *Proceedings of the 4th International Conference on Cognitive Science*, July 13-17, 2003, University of New South Wales, Sydney, Australia. 330-335
- Schriefers, H., & van Kempen (1993). Syntaktische Prozesse bei der Sprachproduktion: Zur Numerus-Kongruenz zwischen Subjekt und Verb. *Sprache und Kognition*, 12, 205-216.
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, 68, B13-B29.