Draft Minutes - UTC #75 & L2 #172
Cupertino, CA - February 24-26, 1998

1. ADMINISTRATIVE ISSUES

1A. UTC Membership Roll Call:
PRESENT: Apple Computer, Inc.; Hewlett-Packard Company; IBM Corporation; Microsoft Corporation; Novell, Inc.; NCR Corporation; Oracle Corporation; The Research Libraries Group, Inc.; Sybase, Inc.; Unisys Corporation; Xerox Corporation. (Total present: 11)

NOT PRESENT: Booz, Allen & Hamilton, Inc.; Digital Equipment Corporation; JustSystem Corporation; Reuters, Ltd.; Silicon Graphics (Total not present: 5)

LIST OF ATTENDEES: See Attachment I

1B. Declaration of Joint Meeting

1C. Approval of Joint meeting Agenda [97-261R]
[#75-M1] Motion: To approve the agenda as amended.
Moved by McGowan, seconded by Umamaheswaran.
Unanimously Approved

1D. Approval of Minutes (postponed) and Review of Action Items [L2/SD2]
Approval of the Minutes and review of action items were deferred.

1F. Review of Meeting Calendar (See Attachment IV)
Review of next UTC meeting date. Problematic for quorum. Apple, NCR, and Xerox not attending.

Action 75-1 for Winkler: Determine whether Unisys meeting facilities are available on May 6-8 and/or May 20-22, by Friday [February 27]

Action 75-2 for Aliprand: Check with other members re their preferred date and dates that they cannot attend by Friday [February 27]. (Dates for consideration: April 20-22, May 6-8 or May 20-22)

1.G. UTC Administrative procedures
1.G.1. Requirements for a complete proposal (Aliprand) [L2/98-040]
Proposals submitted to WG2 require use of the proposal summary form to capture all the information that WG2 needs. There is no mechanism to collect the additional information that Unicode needs (e.g., properties). The WG2 Convenor would be willing to work with us on a synchronized form.

Whistler said that the form should provide sufficient information for encoding a new script, yet not be too daunting to fill out.

Ksar said that the form is there to help submitters. You need guidelines for the user to explain why he needs a script proposal or a set of characters.

Two versions (like 1040 EZ and 1040 forms) were suggested, one for whole script proposals, and the other for proposals for single characters.

Consensus of UTC that Freytag, as Unicode Liaison, take deltas from editorial committee and work with WG2 to develop a common proposal form.

Whistler noted that collation is a new addition to Unicode requirements. The person proposing addition of a script should provide enough information for a default collation.

2. TECHNICAL ISSUES

UTR #8: Version 2.1

Suignard objected to being asked to consider the comments on BiDi when they were sent only one day before the meeting. Davis said that this was due to late receipt of the reference code. Moore pointed out that BiDi was an ongoing discussion issue. Davis said that solutions for edge conditions would not create major disruptions.

Freytag formally proposed Version 2.2 (2 dot) to put the errata together. A discussion on numbering (2.2 vs. 2.1.1 or 2.11) followed. Davis felt that point +1 ("dot release") should be reserved for the introduction of new characters. Freytag thought each dot release will be full delta to the book.

McGowan recommended publication of the 2.1 document, with everything else consigned to a holding bin. Whistler was seriously opposed to any collection that interferes with 3.0. The energy of the editorial committee should be devoted to 3.0. Dacvis pointed out that most software companies have to accept two development streams.

 [#75-M2] Motion: Resolved that the technical content of UTR #8: The Unicode Standard, Version 2.1 is complete, and publication should proceed. All further changes should be placed in a "holding bucket" for review.
Moved by Freytag, seconded by Ksar
10 for: 0 against: 0 abstention; 1 absent (Unisys)

Additional comments regarding 2.1

The inclusion of glyph errata was discussed. Freytag pointed out that we would need to include images of glyphs. Consensus of UTC to include images of glyphs. Lisa Moore recorded specific changes. "OBJ" as the text in the image of the Object Replacement Character (not "ORC") was affirmed.

Ng asked about the relationship between Version 2.1 and the CJK additions which sre coming up. Aliprand explained that 2.1 is to get information out on the Euro and legitimize use of the Object Replacement Character.

[#75-M3] **Motion:** That the UTC accept document L2/98-038, *UTR #8: The Unicode Standard, Version 2.1,* with recommended modifications.
Moved by Freytag, seconded by Whistler.
Approved Unanimously [11]

Freytag commented on how we publish on the web. Striking "status of document" will be substituted by "Standard." He recommended "This version" and "Previous version," but Davis felft there should be only one version.

In addition to an HTML version, we need a ps version also. Whistler said that there also needed to be a printed version with fixed page numbering for reference.

Action Item 75-12 for Moore: Revise L2/98-038 to incorporate agreed modifications (including exclusion of BiDi corrigendum on stack limit [#75-M17]).
Action Item 75-13 for Moore: Apply UTR style guidelines and publish on Web site
Action Item 75-14 for Moore: Announce publication on "unicore" and "news" lists.
Action Item 75-15 for Moore: Arrange for a version of UTR #7 with fixed page numbering to allow for specific referencing.

New fixes proposed for Unicode Datafile - changes to decompositions.

Whistler proposed in document L2/98-060:
- Corrections to decompositions of some Arabic ligature medial forms involving points (use of tatweel instead of space);
- Mark decompositions of parenthesized Latin compatibility forms as explicitly compatibility descompositions, not canonical decompositions (to correct an asymmetry between the decompositions of these forms and partnthesized hangul and ideographic forms).

Whistler: Decompositions are normative in the Unicode standard. For anyone parsing decomposition off the data file, these would be additional changes. We are looking for consistency. If they are using these compatibly, they should decompose accordingly.

Freytag said that there is a meta issue on changing normative properties. We freeze mistakes on characters. We need to have a meta decision of what we do with changes in properties, and we need to do above board. These property changes require a $2/3^{rd}$ majority, but, more importantly, we need to capture the reasoning for the changes.

Edberg said that Apple had to make an exception for these characters.

Ksar asked if there an impact on existing implementations. Whistler replied that there would be, but modifying them for the better.

Ksar also asked whether there would be an input issue for the Arabic characters. Whistler: These 3 characters are inconsistent with other decompositions of medial forms in the standard. The proposed changes are only for decomposition purposes.

Whether to includes these changes in Version 2.1 was discussed. Umamaheswaran asked whether there was a benefit to put the changes in 2.1. Davis replied that there is a real advantage for an implementer to say "conformant to 2.1," but was not aguing that these changes should go into 2.1.

Fixes to UnicodeData file
[#75-M4] <u>Motion</u>: To accept the proposed changes to the UnicodeData file in document 98-060.
Moved by Davis, seconded by Whistler. [11 for, 0 against, 0 abstentions]
Motion approved with required 2/3 majority (10 of 15 members)

Justification for Corrigenda
[#75-M5] <u>Motion</u>: When there is a change in properties approved by the required 2/3 super majority, the justification for making the change must be included. Errata and corrigenda will be published as a numbered collection to facilitate reference. Moved by Freytag, seconded by Davis. Unanimous

Action Item 75-16 for Whistler: Incorporate approved changes, and give justification for changes [#75-M5].

Action Item 75-17 for Freytag: Arrange for numbering of errata and corrigenda on Web site

UTR #9: BIDI reference algorithm

Freytag proposed an Ad Hoc meeting on BiDi, to be held during the week of the WG2 meeting in Seattle. Suignard suggested Wednesday, March 18th after 3 p.m.

Transcoding hint characters

This proposal addresses perfect roundtrip conversion from legacy encoding to Unicode and back. An earlier version was presented in May 1997 (UTC #72 /L2 #169).

Davis: Problem of adding stateful characters to Unicode, and "gumming up" if system is unaware of them. We keep on adding complexity to a plain text standard. There are visible characters and invisible characters. We need statement of consequences in malformed situations; what are the bounds?

Edberg replied that this proposal has no pairing. A transcoding "variant hint" character behaves just like a combining characters, and is not stateful. However, grouping hint

characters are pair-like, and piggy-back on the combining Han structure proposal mechanism. All you lose is round trip variability.

Freytag said that hints will only be interpreted properly in the same domain. Edberg disagreed that these are not useful across platforms.

Davis said that transcoding characters are a limited private use, with a well defined private semantic. To interpret them, you have to know what source platform was, otherwise if I take your stuff it gums up on my transcoding.

McGowan said that the registry, which establishes a meaning in combination, has to be consistent across all platforms. Freytag: If you are trying to transcode into X, you know that, while in Unicode space, the pair has a unique registered meaning. McGowan: When using Unicode from one standard to another, we get into problems.

Whistler said that there are other round trip source problems; for example, MS code page 932 has many duplicates; problem now for Unicode mapping. More seriously, the implication of this proposal is that we are taking on such a registry. This effectively becomes part of the standard and the business of the UTC because these are standards for interchange.

Whistler also asked how this proposal is related to Mongolian and CJK variance. Becker said it is included. Edberg gave a high level view of difference: Mongolian and CJK proposals are specifically for display variance; the variants in this proposal are explicitly only for transcoding. With respect to the registry, the idea is to self–register.

Ksar was concerned that adding 20 encoding positions takes primary space in the BMP. This would create a synchronization problem. The proposal is addressing specific encoding between existing character sets and Unicode.

Ksar felt that it may be easier to add to code points, to make the standard simple and implementable. The proposal is going to cascade into a lot of additional coding. Furthermore, the registry cannot just float: it must also be considered by WG2.

Edberg replied that the proposal is specifically to avoid adding more characters to Unicode; to take off pressure to add variants for round-trip mapping.

Davis felt that the registry would not work, and suggested a UTC compilation of what the hints are developed over time. He asked about the scope: a never ending list or a small number of characters? Edberg said that the number of characters is small, covering round trip compatibility from legacy encoding.

Freytag: One model for regististration would be to use an alternative registry just for logistics, for example, AFII, which has a registrar. To be accepted here, a proposal must have a detailed implementation. Are these variants glyphic or not? Glyphic issues should be addressed in the proposal.

Umamaheswaran said that we need a list of initial characters that are known about. All transcoders in the world will have to be upgraded. Edberg replied that in most cases known about, the characters are already in Unicode.

Whistler said that we could use surrogate mechanism, which gives a single Unicode semantic for the individual character, but do we buy enough? Not clear whether we do. We end up with some transcoding variance, no matter which way we go. We need to see the list. The transcoding hint is easier than Plane 14 tagging.

Davis commented that the transcoding proposal is very like surrogates, using 2 characters. But unlike surrogates, if I throw away transcoding, the initial character remains. The real advantage is that initial character does have well defined meaning (unlike surrogates).

Becker asked whether the variants are a closed set. Edberg said: I have a closed set, but there may be other sets. Aliprand: I have a closed set, variants in the East Asian Character Code. Becker proposed examination of the variants, and make a proposal based upon specifics.

Action Item 75-18 for Edberg: Compile list of Big 5, etc. characters that cannot be transcoded.
Action Item 75-19 for Aliprand: Compile list of EACC characters that cannot be transcoded.

Murray expressed concern about what engineers will do with this, and said that it reinforces myths about Unicode. He likes the Plane 14 proposal, because it makes it difficult.

McGowan asked how lack of this transcoding ability is detrimental to the Unicode Standard? Edberg replied that current mapping tables do not accommodate it. McGowan suggested that the mapping tables should be fixed.

Freytag proposed that the UTC revisit the proposal at the next meting. List of changes, based on the first cut of known characters, is needed for evaluation. If a registry is part of the proposal, it needs to be specified.

Davis: In a list, include a table with two variables: surrogate pairs and hinted characters. Whistler: Since there are a thousand characters, we need to see classes of these things, examples of these.

Transcoding proposal
 [#75-M6] <u>Motion:</u> That the UTC is not yet ready to accept this proposal, and will revisit it at a future meeting. The author should take the following actions in preparing the revision: (1) provide a list of characters that cannot currently be transcoded for evaluaton; (2) narrow the proposal by removing the options; (3) give details about the

registration procedure if this is needed; (4) define the known classes of transcoding problems.
Moved by Freytag, seconded by Roberts.
Approved: 10 for,1 against, 0 abstention

Texin asked, with respect to Big 5 conversion tables, are we relying on their [Apple's] effort? Suignard said that Microsoft doesn't have round trip mapping.

Action Item 75-20 for Edberg: Revise transcoding proposal, according to [#75-M6], and resubmit.

2.C. New Proposals

Formal Criteria for Disunification

Freytag introduced this proposal. There have been a number of efforts to separate existing characters into two, for example, "disunification":of the Euro-currency sign, U+20A0.

Becker said that another question is: what are the property differences; e.g., behavioral differences and context.

Whistler: Is the representation done with different characters with distinct properties or one with ambiguous properties? If there are property differences, are two characters which have distinct properties best or a single character with ambiguous properties? We have plenty of characters in ASCII and Latin 1 that are ambiguous. We handle this by overloading.

Roberts gave the example of "i" with dot over, but Davis pointed out that the casing property is not a normative.

Davis pointed out that we account only for properties that are relevant to applications; e.g., we decided against a decimal separator. The multiple functions of period cause cost and confusion.

Umamaheswaran recommended that criteria for disunification should also be part of 10646 procedures.

Criteria for disunification - Document 98-042, *Formal Criteria for Disunification* [#75-M7] <u>Motion:</u> That the Unicode Consortium submit this document to WG2 as input to the ISO/IEC 10646 procedures document.
Moved by Umamaheswaran, seconded by Freytag.
Approved Unanimously.

Action Item 75-21 for Freytag: Make into draft UTR #11 for April UTC meeting, incorporating examples mentioned at this meeting

Action Item 75-05 for Winkler & Aliprand: Include UTR #11: Criteria for disunification on agenda of UTC #76/L2 #173 joint meeting.

Action Item 75-05 for Aliprand & Winkler: Include on agenda for UTC #76/L2 #173 joint meeting the format of electronically distributed data (Winkler/Freytag).

"Width" Character Property

Freytag introduced this proposal. Some East Asian character sets combine typographical attributes and encoding. There are a number of duplicates in some Asian chracter sets. The four categories of widths that occur are: "Half-Width"; Ideograph (equivalent to Wide); "Full Width"; and ambiguous (no narrow form in legacy).

We need to know a character's width, because the wide and narrow characters have different behavior. There are two basic inherent classes: Narrow and Wide.
- Narrow: "Half-Width" and Neutral
- Wide: "Full Width," Ideographs, Hangul, and Ideographic punctuation

It is worthwhile in practice to pull out half-width. Characters are from non-western scripts and are a mixed bag, including half, narrow and ambiguous. Half-width is defined by content of legacy character set. Addition of single byte characters is severely curtailed; limited to 96 characters or 130.

The important points to realize are that the half-width is a superset; and, ambiguous is based on best current knowledge. Superset of all characters - in wide characters - of legacy character set. In a practical application, need a quick way to sort.

Roberts: I have an implementation that partitions Unicode as neutral, ambiguous and wide. I cannot make this work. I want to distinguish neutrals and narrows. Freytag said that we are not trying to create something that covers each and every part of Unicode. The least leveraged is when whole blocks can be used.

Ksar asked how addition of a new property will impact the Unicode database. Davis said that it would require addition of another column

McGowan asked whether this would be an informative or normative property? Freytag replied that he had no strong opinions on it, and would not be opposed to a two-step process: informative first and then normative later. He pointed out that the UTC doesn't invent properties; we discover properties. Davis also favored informative initially.

Whistler pointed out that we are setting precedent for the database. Will be adding to the end? McGowan said "as long as we don't put fields in the middle."

Whistler said that we already have a mismatch of data files (database and properties file). Freytag proposed an action item for the formatting of electronically submitted data.

Texin liked Freytag's analogy of discovery, and would like to see list of possible properties that are being considered. He added: I would like to make a motion that we maintain a list, properties to do with punctuation.

Freytag replied that there is a misunderstanding. Our discovery has shown these to be orthogonal. Wide characters are always Asian characters. You could argue that we already have half-width characters. The property is already orthogonal.

Whistler made some specific proposals. First, in turning into a draft technical report, tables should be sequential tables. Terminology, as stated, is very confusing, and "Narrow" in table should be changed to something else. I would like to have a specific definition. "Narrow is…" Davis added: Define as "unmarked" rather than "neutral." Where there are seeming inconsistencies and compatibility characters we need to clarify. Texin recommended use of "East Asian" e.g., EA-full width. East Asian. EA- ambiguous East Asian. EA - half-width East Asian. B - not East Asian.

Action Item 75-62for Freytag: Make into draft UTR for April UTC meeting, incorporating examples mentioned at this meeting.

Changes in Property Assignments for Quotation Characters

Freytag: I want to distinguish this paper and the proposal. We are having some difficulties in dealing with quotations. We appear to have been too eager in assigning opening and closing properties. We need useable heuristics for specific scripts regarding these properties.

All punctuation marks are given the following:
Ps (punctuation start)
Pe (punctuation end)

In Czech, and German, the beginning quote is like an English ending quote and is dropped down to base line.

Guillemets are used differently in different languages, e.g., >>Svenk>> and
<< French >>

The quotation marks in Japanese vary whether they are opening marks or closing at the beginning and end, and whether above or at the baseline.

What appears to be "normal" is script specific. We need to change the term "preferred" in Unicode annotations.

Action Item 75-22 for Becker: Check Japanese standard(s) to determine position of Japanese quote marks relative to "cell" boundaries, and send findings to Freytag.

Action Item 75-23 for McGowan: Check Japanese typographical manuals re quotation marks and send findings to Freytag.

Freytag said that he has a number of errata to file - adding to the bucket – with a new proposal to change the starting and ending of quotation marks. Propose: Quotation "often" starting, Quotation "often" ending, that is, context dependent. In addition, language that has to do with what are usable heuristics for analysis.

Whistler pointed out that punctuation values are not normative. We need further sub-division to a category.

[#75-M8] <u>**Motion:**</u> That the UTC adopt document 98-053, *Analysis and Proposed Changes in General Category Property Assignments for Quotation Characters*, as amended with the comments from the discussion.
Moved by Davis; seconded by Sargent. 10 for, 1 against, 0 abstention
Motion approved with required 2/3 majority (10 of 15 members)

Whistler noted that implementation of these changes entails clarification of special case errata. Changes need to be made in:
1) book in several places - block and chapter
2) Unicode Database
3) Unicode READ ME file
4) Properties file

Becker, on behalf of the UTC, thanked Freytag for opening this can of worms. Freytag replied that the next proposal will deal with dashes. He thanked Suignmard for his valuable input on this proposal.

Action Item 75-24 for Freytag: Incorporate comments from discussion and other feedback. File as corrigendum and give justification for changes.

<u>Character Properties for Control Codes</u>

Ksar said that this question arose regarding 10646 and the usage of "new line" character. He was under the impression that ASCII and ISO 646 have been modified to deprecate "new line."

Umamaheswaran said that the problem comes about because C has violated ISO 646.

Ksar recommended that Version 3.0 include a clarification in order to avoid any conflict like this.

Whistler said that some properties for control characters are already defined. Clarifying the knotty problem of line breaking behavior is on our "to do" list, and text is being written.

Davis became aware that there was a problem when they used mapping tables on Unicode site for control characters, and they mapped to unknown. Control characters are not in the mapping tables; it behooves us to fix this.

Action Item 75-25 for Davis: Begin process to add control characters to mapping tables.

Umamaheswaran asked about ISO 10646: Does ISO 10646 know about the C world? What should the Unicode book say? Does C violate ISO 646? The C standard is in violation of 10646 and Unicode. This is an L2 thing, since C is an L2 issue.

Action Item 75-02 for Winkler: Add C violation of ISO 646 to agenda of L2 meeting.

McGowan requested that all comments about clarification to text in the book be sent to him. Ksar thanked McGowan for taking on responsibility, and asked about the deadline for this. Moore said that it formed part of the editorial committeee work.

Motion to Adjourn for the day [Wednesday, February 24] at 5:00 p.m. Adjourned by Consensus.

Thursday, February 25 Meeting called to order at 9:10 p.m.
[Apple, HP, IBM, Microsoft, NCR, Novell, RLG, Sybase, Unisys (9 members; quorum is 8). Xerox at 9:30 am and Oracle at 9:40 am]

UTR#10: Unicode Collation Algorithm

Davis gave an overview. Collation elements have three weights: levels 1, 2, and 3. The tailoring syntax allows a number of different operations because you need to do different things. [Goals are stated on first page of document]. Page 5 has rearrangement; page 9 has new operators; page 9 and 10 new explanatory material. There are some typos on page 12. Within document, table format is detailed. FCD 14651 had weight that was not equal to Unicode. Specials - can distinguish on tertiary level. Problem of 4th level is implementation cost.

What do you do if "Ä" is encoded as a combining sequence? Weights at second level indicate what it should be if not handled. But if combining sequences are handled, you use equivalence. The difference is that you don't get exactly equivalent positioning if you don't do the normalization. You can do either way for normalization. One big switch: Do you want to do normalization or not? Data tables, as are, support Latin 1, Latin 2, etc.

Umamaheswaran asked does the proposal support collation of French accents?
Davis: Not easily. Normally, you register changes in letters going forward. This is a problem in multilingual text., since you do not want to apply French procedure to all the text. There is a solution, but it is a bit messy to implement.

Umamaheswaran: Does this agree with the tables in 14651? Davis said that, functionally, the proposal meets 14651.

Whistler said that the tailorable 4th level of 14651 cannot be accomodated, unless 14651 allows conformance without the tailorable 4th level. The problem is their **requiring** the 4th level. The way that the Unicode algorithm is set up is that everything is given 1st, 2nd, and 3rd level. Tailor to turn off if needed to default and ignore first three levels. Differences of punctuation is resolved at the "last" level. You cannot specify a fourth level.

Umamaheswaran: We have to prove to them that the fourth level is not required.

Whistler: The Unicode collation algorithm reproduces all the same Canadian benchmark examples correctly. Three levels can be contained within 32 characters. The 4th level goes beyond this limit, which imposes a significant cost for implementations (space cost not money cost).

Davis: We've asked many times for real world examples to illustrate (for example) the statement that Swedish requires 4 levels, but no "real" language cases have been demonstrated. If you require a fourth level it all balloons and cost escalates. This Unicode algorithm can collate all the real life words of the world.

Freytag pointed out that there is an associated cost. If memory is free then there is no cost. Space costs. Implementation costs.

Whistler: Everything that I have been given in Swedish can be done. The 4th level justification seems to be based on empty claims; combined with a claim that special order is needed for resolving at the last level.

Ksar: This is a very good document to describe. My concerns: Do you want to combine 66 with this? As Umamaheswaran mentioned, show example that it is backward compatible. Suggests combining 066 as informative annex. State that Canadian benchmarks have been tested. As far as the Swedish case, keep the door open whether it works or not. As far as this document in Unicode 3.0, will it be informative or normative? We want uniform and consistent.

Davis: The goal is to say "If you are implementing the Unicode algorithm, then you have to conform to these requirements, i.e., three levels only."

McGowan: It is important to reproduce the same results given the same data.

Texin: With respect to implementation cost and levels: How is it applied regarding Asian characters? How do we apply it in a limited environment where more than 1 or 2 levels are not required? Is this case conformant? Whistler replied that you can support three levels and allow certain programs to ignore some levels.

Aliprand noted that the Royal Library of Sweden is a Unicode member who can provide proof of whether the fourth level is needed for Swedish.

Action Item 75-26 for Aliprand: Check with Royal Library of Sweden to see if 4th level is required for Swedish.

Umamaheswaran expressed concern that ISO goes one way and Unicode goes another, wants one standard not two different ones. There is a need to conform to the Canadian collation standard which has a 4th level.

Davis replied that had we not been involved with 14651, it would have been unimplementable. We found massive syntax errors and it doesn't explain how the sorting syntax is to work. The thing we would like to see is for 14651 to relax conformance requirements, so that Unicode Collation Algorithm can be conformant. It is hard to demonstrate the non-existence of things.

Action Item 75-27 for Umamaheswaran: Send Red Book examples to Whistler and Davis.

Winkler said there is a rumor that polytonic Greek needs seven levels. Whistler said that it works fine with three levels. We need real examples to test such assertions, but those examples have not been supplied. Davis added that they have to be real life examples; otherwise, it is very arbitrary.

Winkler asked about space handling in German. Davis replied that basically you can turn spaces on or off, and space is defined as ignorable by default.

Winkler said that 14651 has requirement of handling upper case and lower case, and asked whether case is switchable. Whistler said you have to define a parameter to do collation to do one way or another. Default table has case distinction in the tertiary level.

Winkler: I hope that an amended copy [Unicode Collation Algorithm] is available for me to present to WG29 in June. Also, with comments.

Freytag asked whether any of the technology in the Unicode Collation Algorithm is under patent. Davis replied: To my knowledge, no. Freytag asked because he had heard that IBM had a patent at the 4th level. Whistler said that this would present an interesting problem for FCD 14651 if the 4th level is patented. Moore said that IBM has 3 patents, awarded in '91 and '92.

Action Item 75-28 for Umamaheswaran.: Investigate IBM's patents re collation algorithm. Provide formal clearance from IBM re patents.

Roberts noted that a four level API is very common in industry. He expressed concern about competing standards and some new default compatabilities which are not in a

fourth level. As an example of more levels, he said that a Japanese example of a phone book has 7 levels.

Whistler responded that telephone books have special requirements, which are out of scope of this default collation algorithm. More levels deal with specials across different character sets. Phone books depend upon what collation one wants to use.

Roberts mentioned choice of APIs. Whister and Davis both emphasised that this is not an API. APIs that have been implemented do not look like this. Roberts asked: Do you expect APIs to implement this? Whistler said they could be implemented on top of tables that look like this. At Sybase and with Java, we have implemented default tables that collate on this.

Murray asked for clarification on Japanese. Matching of hiragana and Katakana? Width issues? Whistler responded that the default table for Japanese is based on the Japanese comments on 14651. Width is treated as tertiary variance.

Yu asked how to achieve culturally correct sorting for CJK if this was a primary level assignment. This would be achieved by tailoring.

Davis said that the reason for the 4th level is that a number of people want the first 3 levels not to change. One could say, my customers don't care about the fourth level. It is possible to be compliant and ignore the 4th level.

Whistler said that the point is to produce the same results. If you implemented with 17 levels and can produce same output, you would be considered conformant.

The Chair expressed awe of the amount of work that went into this proposal. [Applause]

Suignard and Freytag made specific comments, including a need for explicit definitions and scope of the Collation Algorithm.

Winkler mentioned other sorting applications: TC37 (Bibliographic sort); TC 46 (work item for European character sets); CEN (work items on mulitlingual sorting).

Action Item 75-29 for Aliprand: Send NISO sorting examples to Whistler and Davis

Action Item 75-30 for Winkler: Send TC46 and TC37 sorting documents to Whistler and Aliprand.

Hart formulated some formal input to L2:
1) Objective is to have one standard in this area.
2) Explicit statement of producing same results over details of implementation.
3) Examples of where the Unicode Algorithm works or doesn't work in relation to 14651 and other standards. Maybe we need to subdivide.

4) Desirable to have tables in machine-readable form. It is of limited value to have on paper.

Moore said that there may be a problem for EBCDIC systems. If we try to reproduce, this sort of table will not work easily for us. Davis suggested tailoring. Whistler identified the 4th level as the problem. 4th level is resolved by Unicode value, but the 4th level in EBCDIC was resolved by EBCDIC order. Could be tailored at 3rd level to get close.

Umamaheswaran: We have the issue of legacy data. Davis responded that the existing algorithm could be used for legacy data. Moore explained that at a lot of our 390 systems, they want to use Unicode. There is a desire for Unicode collation tables, but there may also be requirement for no change from the current situation.

Umamaheswaran said that the 4th level is fixed. The requirement is to reproduce what I have today. Whistler: As stated now, the 4th level cannot be tailored. We see no reason to introduce a 4th level. Implementation of Java cannot be tailored. IBM is a case of interoperability. Need to introduce a tie breaker.

Ksar: Somehow I feel this 4th level thing is becoming more important because of interoperability. Could the 4th level be optional? What is the cost?

Whistler: I don't see any reason that it could not be an option: tailorable 4th level or not. There would be a choice up front to declare "not tailorable."

Action Item 75-31 for Umamaheswaran/Moore: Research scope of problem for use of proposed collation support tables with EBCDIC systems.

Davis: What if we had three options: 1) Tailorable 4th level; 2) Fixed 4th level (Unicode); 3) no fourth level, essentially wrapping 4th level into 3rd level?

Becker: Are you proposing to allow tailoring of the 4th level by code set?

Davis: Option 1= fixed 4th level (=Unicode); Option 2 = tailorable 4th level (after you weight everything on the first and through third level); Option 3 = combine 3rd & 4th level. There is a big difference between combine and discard.

Whistler: Discard is a special case for what you do with searching. Matching should be able to disregard all lower levels successively.

Option	Sort Key	Bit Requirement
Option 1 (Unicode)	Approx. + 50%	32
Option 2 (tailored)	Approx. + 50%	>32
Option 3		?32

(combined 3rd and 4th level)		

Freytag asked why do we need to specify option 3 in our TR? Davis replied: For reproducibility across systems, to give a minimal level of conformance.

Whistler said that Level 3 was equivalent to Option 1 + tailorability into 3rd level. Hart agreed, saying that Option 2 is more powerful than 1 and more powerful than 3. He suggested eliminating Option 3.

There was a question about sorting of surrogates. Davis replied that you can place a pair of Unicode values anywhere in order. Can use surrogates with 32 bit keys. Freytag asked whether this was an efficient way to deal with 250,000 characters in other planes. Davis said that the key is to map a set of characters to set of collation elements.

Freytag asked about collation of Scandinavian surnames. His example was discussed. The outcome was that Davis agreed to add "add normalization" as part of Option 1 conformance.

Texin: With respect to interoperability – what is meant? If it is the ability to produce tables that are being produced elsewhere, there is no provision for signature mechanisms. Whistler responded that the problem of named collations of particular types beyond scope of standard.

Whistler appreciated the tremendous feedback. Davis noted that a default set of parameters needs to be defined, and suggested going public with the next draft.

[#75-M9] <u>Motion:</u> The UTC directs the authors to incorporate feedback into document 98-065, *Unicode Collation Algorithm*, draft dated 98.02.25, to post it for public review, and to solicit public feedback.
Moved by Davis, seconded by McGowan
10 for, 0 against, 1 abstention (Novell) [Xerox out of room]

Action Item 76-32 for Davis/Whistler: Incorporate feedback and post draft UTR for public review.
Action Item 76-33 for Davis/Whistler: Solicit feedback on draft UTR from outside UTC.

5. W3C Issues

New RFCs

Winkler added one more new RFC: RFC 2277 – FTP internationalization. Internet Draft 959 & 1123.

Action Item 75-63 for Winkler: Post URL for RFC 2277 draft to "unicore."

RFC 2244: ACAP – Application Configuration Access Protocol

Moore: I sent a letter in which I basically said it was not good to endorse standards that use non-comformant UTF-8. Becker asked whether there were notes from the IAB Character Set meeting. Moors said that she had some notes.

Freytag said that we can find a suitable place on our Web site to document violation of UTF-8.

[#75-M10] <u>Motion:</u> The UTC believes that the statement in section 9 of document *RFC 2244: ACAP – Application Configuration Access Protocol* (as reproduced in document 98-063), that "no clear consensus on a single method is apparent at this stage" is incorrect. Furthermore, the Multi-Lingual String Format violates UTF-8.
Moved by McGowan, seconded by ???
Unanimous Approval

Action Item 75-34 for McGowan: Publish statement based on [#75-M10] on Web site

Action Item 75-35 for McGowan: Communicate with IETF re problems identified in #75-M10.

Action Item 75-36 for Freytag: Evaluate FTP internationalization Internet Draft

4. SPECIFIC SCRIPTS

Western Music

McGowan introduced the revised proposal for Western Music. Corrections are to be made to the text. The issues raised have been answered.
1) BiDi: Mike Ksar can verify Western music in Arabic is left-to-right. Western music in Hebrew is also left-to-right. Thus, there are no BiDi implications for this proposal.
2) start and stop have been explained;
3) unification/disunification, particularly dots, have been explained, e.g., p. 2. Staccato dot (distinguished from Latin character over dot);
4) identification of reserved positions are in names list.
The character names list is corrected. Character properties have been added. Page 9 shows examples of canonical decomposition.

Moore asked if Western mustical symbols are all strong left-to-right characters. McGowan confirmed this.

McGowan noted the discussion of examples on page 9, and of Baroque ornamentation on page 10. The Baroque ornamentation examples come from Bach's score that he wrote for his children.

Roberts: Beginning and ending. If you don't have an end tie, implementer is free to interpret.

Whistler: Most people aren't going to be happy about. Thus BiDi implies a change in sharps, flats and naturals – left-to-right.

Aliprand: Characters need to be consistent.

McGowan: Make them all neutral.

Becker: If we have a problem we have it now.

Freytag: We need to make sure that we don't have another problem related to BiDi.

Moore: Strong left-to-right…

McGowan: I think they all should be strong left to right.

Freytag: I don't know if we are doing it right.

Sargent: Nicer if we don't add more neutrals.

Freytag: Murray's point about neutrals slowing things down…

Rick: I have a book in Arabic and Hebrew about Western Music. Sharp shouldn't migrate, it is a word.

Action Item [ ] for Sargent, Becker, Roberts, Texin: To find out if we have issues with the Western music proposal, in regards to BiDi.

McGowan: In the absence of examples to the contrary, we should go left to right.

Freytag: The proposed properties are strong left to right.

Western music proposal
[#75-M11] <u>Motion:</u> To made the editorial changes noted at this meeting to document 98-045, *Proposal for Encoding Western Music Symbols in ISO/IEC 10646*, and to then submit the proposal to WG2 as a contribution to ISO/IEC 10646.
Moved by McGowan, seconded by Umamaheswaran
Unanimous Approval

Action Item [ ] for McGowan: Work with Perry Roland on final version of Western music proposal [#75-M11].

Action Item [ ] for McGowan: Supply formatted table entry for Western music proposal to Freytag for web.

Texin: Is order of chart, the order they will sort?

McGowan: Irrelevant. Binary order is fine.

4.B. Ideographic additions
4.C. Additional non-ideographic characters

4.C.1. Enclosing triangle (McGowan) [L2/98-056]
McGowan: Combining upward pointing overlay. Overlays have properties of being a base.

Enclosing triangle
[#75-M12] <u>Motion:</u> That the UTC accepts document 98-056, *Proposed Triangular Overlay Character*. To submit the proposal to WG2 as a contribution to ISO/IEC 10646.
Moved by McGowan, seconded by Carroll
Unanimous Approval

Action Item [ ] for McGowan: Supply formatted table entry for Enclosing triangle to Freytag.

4.C.2. KIP sign: Laotian currency sign (Umamaheswaran) [L2/98-061]
Freytag: KIP sign is relatively straight forward.

Umamaheswaran: At moment, we don't have glyph.

KIP currency sign
[#75-M13] <u>Motion:</u> That the UTC accepts document 98-061, *KIP Sign – Laotian Currency Sign.*
To submit the proposal to WG2 as a contribution to ISO/IEC 10646.
Moved by Umamaheswaran, seconded by Roberts
Approved by consensus.

Whistler: We should suggest code points. Recommend to encode at 20E4.

Michel: … pipeline characters 20E2 and 20E3.

Whistler: We have changed our procedures somewhat.

Aliprand: Should be part of the UTC procedures.

Action Item [   ] for Umamaheswaran: Supply formatted table entry for KIP currency sign to Freytag.

4.C.3. LATIN CAPITAL and SMALL LETTER H WITH CARON for Romani language of Finland (Winkler) [L2/98-062]

Aliprand: The standard line is that you can make with [Joan, I missed this?????] How opposed are we to this?

Freytag: Advocate whole Latin caroning!

Whistler: I would suggest that you don't do that.

Freytag: Be careful not to alienate but don't include a principle.

McGowan: You say give it a code point, but we already have a means to do that.

Aliprand: This comes from the Finnish body.

McGowan: We will have a thousand of these.

Freytag: Combining characters are out there. Collectively our membership has not done what is necessary to support.

Sargent: We will in a couple of years.

Whistler: I see the validity of Asmus' claim. We have been caving on this for years & we will continue to cave for years. The incremental cost is real, but the political issue is rather small. It won't get into an implementation for years, either way. Doesn't take too much to "never mind either way."

Suignard: Very strong pressure in Europe. The Nordic countries are good about getting these things done.

Freytag: The Nordic countries have the highest computer literacy in the world and active governments. The political philosophy is to be culturally inclusive.

Aliprand : Politically speaking-> 1) the Nordic countries are good allies, not alienated; and, 2) it is going to be approved by WG2 anyway.

H, h with caron
 [#75-M14] <u>**Motion:**</u> To disapprove the addition of the two characters in document 98-062, *Proposal to add 2 Latin characters to ISO/IEC 10646*, to the Unicode Standard because they can both be encoded using the combining caron on a base letter.
Moved by McGowan, seconded by Honomichl
5 for, 6 against, 0 abstention
Motion failed

[#75-M15] <u>**Motion:**</u> To accept the two characters in document 98-062, *Proposal to add 2 Latin characters to ISO/IEC 10646*, for addition to the Unicode Standard.
Moved by Moore, seconded by ??? (Freytag?)
6 for, 5 against, 0 abstention

Motion carried

Action Item [ ] for Whistler: Add Proposal to add 2 Latin characters to ISO/IEC 10646 to Pipeline documents.

Action Item [ ] for Moore: Supply formatted table entry for H, h with caron to Freytag

4.C.4. Two ecological symbols (Everson) [L2/98-025]
http://www.dkung.dk/jtc1/wg2/docs/n1661.htm
Aliprand: We need to provide feedback to Michael Everson.

Freytag: One is a recycling symbol and the other is *Der Gruen Punket* (German symbol of a consortium. You may use the symbol, if you are a manufacturer and you are required to take back packaging material). This is an incorrect coding of someone's private symbol. It is owned by someone.

Whistler: Let us instruct our Liaison that, 1) it is a commercial label; 2) it is printed in color (and the color has meaning. It is clear that these are not dingbats.

Action Item [ ] for Aliprand: Give appropriate feedback to Michael Everson that proposal lacks due diligence.

4.C.5. Support for implementing inline and interlinear annotations (Freytag) [L2/98-055]
ftp://www.unicode.org/WorkingGroups/EastAsia/Ruby.htmlRuby (Postponed presentation. To follow Syriac proposal on Friday)

3. SPECIFIC SCRIPTS (return to script discussions)
3.A. Whole Scripts

3.A.2. New Proposals for BMP

a. Syriac script (Kiraz, Hasso & Nelson) [L2/98-050, L2/98-051]
http://www.cl.cam.ac.uk/users/gk105/syrcom/unicode/
Syriac script examples [L2/98-052] available upon request.
⇒ This proposal will be discussed on Friday morning.
b. Coptic (Everson) [L2/98-022]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1658.htm
http://www.indigo.ie/egt/standards/cy/coptic.html

McGowan: Coptic is controversial and disunifies.

Hart: Glyph same for upper and lower case.

Aliprand: That should be explained in proposal.

c. Tibetan Extensions (CN/GB/IE proposal) [L2/98-024]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1660.htm

McGowan: On the Tibetan Extensions, this was pulled out as controversial. We have seen it before. We have no evidence of due diligence. Lee Collins and Apple would like to reserve four columns for Tibetan Extensions. Nothing has been demonstrated that control characters are needed. If he would split apart into piece that is non-controversial, and a part that is controversial. Could get. OFD is in total controversial. FF is contradictory. Most characters controversial. Others are very controversial.

Suignard: No supporting examples were submitted.

McGowan: We have no examples and no justification.

Action Item [ ] for McGowan: Provide detailed critique to Freytag.

d. Glagolitic (Everson) [L2/98-023]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1659.htm

McGowan: UTR #3. Similar to Joe's UTR. Has six more letters.

Becker: Everson should have cycled to Cyrillic scholars. There is no evidence of contacting scholars.

Aliprand: We need to use the model of Syriac proposal [for its thoroughness and completeness]. We need to circulate and endorse this model of proposals.

Becker: We have no pull from people who want it.

Aliprand: There is some pull from British library.

Becker: Whoever is submitting, we have no procedure for proof of usability.

Freytag: We need to treat as a serious submission and write letter with recommended changes.

Action Item [   ] for Becker: Evaluate 98-023 (Glagolitic proposal)

e. Buginese (Everson) [L2/98-021]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1657.htm

Buginese is acceptable. Only known implementation orders in way.

3.A.3. New Proposals for Plane 1: Current scripts

a. Avestan (Everson) [L2/98-031]

http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1684/n1684.htm

McGowan: I had a conversation with an expert. Fine to encode, but cannot be unified with a Palelavi (sp?

3.A.4. New Proposals for Plane 1: Historic scripts (Everson)

a. Egyptian hieroglyphs [L2/97-266, 267]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1636/n1636.htm
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1637/n1637.htm

McGowan: Due diligence has been done. (Waited a long time to get software). This is ready for "Prime Time"! No reason to diverge from scholarly community.

Roberts: What about beginning and ending cartouche.

McGowan: It is fine as it is.

Freytag: We don't want to diverge from what Egyptologists are doing but require it be checked with existing implementation.

b. Meroitic [L2/97-268]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1638/n1638.htm

Freytag: We need to formally note that we do not want parenthetical remarks. Unicode is writing directions and properties. Provide properties and writing direction information if missing.

c. Old Persian Cuneiform [L2/97-269]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1639/n1639.htm
d. Ugaritic Cuneiform [L2/97-270]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1640/n1640.htm

McGowan: All of these proposals need documentation of support from the user community.

3.A.5. New Proposals for Plane 1: Fictional scripts (Everson)
a. Tengwar [L2/97-271]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1641/n1641.htm
b. Cirth [L2/97-272]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1642/n1642.htm
c. Klingon [L2/97- 273]
http://www.dkuug.dk/jtc1/sc2/wg2/docs/n1643/n1643.htm

Freytag: There are living scripts that we must deal with before we deal with fictional ones.

Umamaheswaran.: What if there is a clear commercial need?

Freytag: First, living scripts; second, commercial/historical; and then 3) fictional.

Whistler: I don't want to shut the door.

3.A.6. Informational: Coming Proposals
a. Burmese (Collins) [l2/98-044]
http://www.unicode.org/pending/burmese/Burmese.html
b. Additional proposals (Everson) [l2/98-032, etc.]
"unicore" e-mail from Michael Everson dated Sun 15 Feb 1998, Subject: Yet more new documents.

Action Item [ ] for Aliprand: Provide feedback to Michael Everson (ccWG2) on the proposals that were discussed at this meeting.

3.D.1. Javanese (McGowan) [L2/98-041]
McGowan: This is for information only. Please read it.

Freytag: We should officially respond to him.

Suignard: This is an example of this being a good proposal.

Action Item [   ] for Aliprand: Inform Jerome Hellingman that his Javanese proposal was distributed to the UTC.

3.D.2. Bangla (Aliprand) [L2/97-287 & L2/98-058]
Request to incorporate Bangla (Bangladesh coded character set) in 10646 [L2/97-287] & draft response: "unicore" e-mail from Michael Everson dated Friday, 30 January 1998, Subject: WG2 N1634 (Bengali)

Freytag: I said I could support this in WG2. I was happy to do so.

Aliprand: We need official response.

Bangla
[#75-M16] <u>Motion:</u> That the Unicode Consortium submit a response to SC2/WG2 N1634, *Incorporation of Bangla (Bangladesh) coded character set in 10646*, considering as input for the response the points raised by Michael Everson in his draft response (document 98-056) plus the comments on it from Asmus Freytag.
Moved by Freytag, seconded by Winkler
Unanimous Approval

Action Item [ ] for Aliprand: Draft a UTC response to document SC2/WG2 N1634, per [#75-M16]. Circulate to members for feedback; incorporate feedback, submit to WG2 via Unicode liaison.

Action Item [ ] for Oesterle: Send free copy of Version 2.0 to Bangladesh Standards and Testing Institution (get address from 97-287).

Return to earlier Agenda Item:
1.D.Approval of minutes of Joint Meting UTC#74 & L2 #171 [L2/98-039]

Approval of minutes
[#75-M17] <u>**Motion:**</u> To approve the Minutes of UTC#74/L2#171 joint meeting (document 98-039) as amended.
Moved by Moore, seconded by Winkler
Unanimous Approval

[Joan – this is an awkward transition]
BiDi corrigenda and UTR #7 (Version 2.1)
 [#75-M18] <u>**Motion:**</u> That the corrigendum to the BiDi algorithm dealing with exceeding the stack limit not be included in Version 2.1 because the text specifying the corrigendum has not been reviewed by the UTC.
Moved by Moore, seconded by Roberts
Unanimous Approval

Action Item [ ] for Aliprand/Winkler: Contact Jenkins re status of AI 169-09 (May 97: provide updated proposals for all scripts for submission to WG2).

Action Item [ ] for Whistler/Freytag: Investigate completion of 72-26 (May 97: Apostrophe – character properties).

Action Item [ ] for Hart: Supply text of PDTR for Character/Glyph Model to Freytag (so 169-56 can be completed)

Motion to Adjourn for the day [Thursday, February 24, 1998 at 5:00 p.m.]
Adjourn by Consensus.

Meeting Convened Friday, February 25, 1998 Meeting called to order at 9:00 p.m.
[Members 10, quorum is 8]

Action Item [ ] for Aliprand/Winkler: Check with Goldsmith re status of 73-57 ff. (IANA charset registration).

Action Item [ ] for Aliprand/Winkler: Transfer AI 73-55 (Editorial issue) to Julie Allen

[Return to previously postponed presentation]

3.A.2.a. SPECIFIC SCRIPTS – Whole Scripts – New Proposals for BMP

George Kiraz: Collapsing of various projects on Syriac
Paul Nelson: Unification of shapes.

Nelson: Note attachment of errata sheet. Syriac is not part of Arabic block. Same handling as Hebrew (European numbers, BiDi). Question of do we add a character…

Becker: The goal is to not duplicate if at all possible.

Freytag: We do lay out numbers differently.

Ksar: BiDi algorithm takes care of property, same as Arabic North African.

McGowan: Any revision of BiDi needs to be mentioned in book, "it behaves the same as…" in describing the BiDi characteristics.

Moore: The algorithm takes care of this…

Nelson: The Syriac abbreviation mark, control character -> to run to end of area. Not a formatting type of mark. As far as I know, it is used only in Syriac.

[Joan, to expedite minutes, I left out more of this discussion. Will finish later. More….]

Action Item [ ] McGowan/Oesterle: Add George Kiraz to "unicore" distribution list.

Ksar: What about fonts for Syriac?

Kiraz: Syriac fonts today include bold, italic, point sizing. East Syriac is boldish by nature

Ksar: Mixed Arabic and Syriac text?

Kiraz: West Syriac and Estrangela are fine; East Syriac looks boldish.

Syriac proposal
[#75-M19] <u>Motion:</u> That the UTC accepts, in principle, the proposal for Syriac script specified in document 98-050, *Proposal to encode Syriac in ISO/IEC 10646*, together with the errata and action item working documents. To submit a revised proposal, incorporating the errata and action items, to WG2 as a contribution to ISO/IEC 10646. Moved by Sargent, seconded by Ksar
Unanimous Approval

The UTC Chair notes that this is an excellent model of a well-organized and complete initial proposal. Can serve as a standard.

Action Item [ ] for Aliprand/Winkler: Include on agenda of UTC #76/L2 #173 joint meeting:
Properties for Syriac characters (Kiraz/Allen).

Action Item [ ] for Kiraz/Nelson: Prepare final version of Syriac proposal, incorporating errata and action items [#75-M19]

Action Item [ ] for Kiraz/Nelson: Prepare separate proposal for Eastern and Western Syriac crosses as miscellaneous symbols.

Action Item [ ] for Kiraz/Nelson: Provide draft properties for Syriac script characters, and for proposed additions to the Arabic block (non-spacing hamzah,, etc.) and Miscellaneous symbols block (Syriac crosses).

Action Item [ ] for McGowan: Supply formatted table entry for Syriac proposal to Freytag.

Action Item [ ] for Aliprand/Winkler: Serve as UTC contact point on Syriac proposal.

[Return to earlier agenda item]
3.C.5. Support for implementing inline and interlinear annotations (Freytag) [L2/98-055]
ftp://www.unicode.org/WorkingGroups/EastAsia/Ruby.htmlRuby

Freytag: We discussed this in August. In paper there are four common forms of things not straight linear text. I'm aware of three implementations and all three are using different separator and terminator. You can have a very cheap feedback …

Whistler: Have you thought through the consequences of arbitrary shuffling of these?

Freytag: Yes, but not in written proposal.
1. You terminate at nearest block or line separator.
2. From fallback, you can next from the fallback
3. If you get out of order [Joan, I missed this …????
4. The existing programs in Japanese have a limited amount to be handled in Ruby.

Texin: What is the impact of handling, and secondly, is it specific to Japanese or does it work for other applications?

Freytag: It is not tied to any type of character. Off the shelf implementations in Japan are pedestrian. There is a Ruby input area.

Texin: For question #2, any system that is going to support it will need to be defined.

Freytag: Ruby is British, meaning small font.

Becker: Ruby is one interpretation?

Freytag: Intent is that it is interlinear.

Becker: Do we want to open or close door to infinite ways of looking?

Chris: Option of all of these Ruby… where to place Ruby. "Rubi" is a Japanese word now. The same use as Ruby.

Whistler: In principle, you don't want to create bad behavior if you embed. What happens if I put an annotation on a Ruby? On nesting of Ruby, are there practical limits?

Freytag: Depends upon how recursive formatting library is. However, 80% of editing is by dialog. Select and annotate. For the purpose of the proposal, conformant application doesn't have to do anything.

Suignard: The selection model becomes very complex. Limitation to one line. Otherwise get into word-breaking issues. Its coming in maybe two years. A conformant application doesn't have to do anything.

Texin: I'm concerned about defining nesting levels, with scope in context of Ruby.

Freytag: Ruby is part of everyday Japanese. This is the fourth time this has come up for discussion.

Texin: This is very close to superscript.

Suignard: Difference between superscripting and Ruby is that Ruby can be removed without affecting base content.

Sargent: Have found application for mathematics.

Roberts: Other uses?

Freytag: Aim is to make solution for Japanese and Asian languages. I am going to make a feature that works well for Japanese and Chinese; if it works for other uses, great.

Umamaheswaran.: Have you thought about BiDi controls?

Freytag: Right to left and left to right. Arabic embedding levels may be a problem.

Texin: I want to understand the behavior. If I want to cut and paste with Ruby, what happens with annotation?

Freytag: The current state is that you get all. Implementation specific.

Sargent: I don't usually think of this as part of the user interface. Can go into select whole object mode.

Roberts: If you have large Ruby strings, do they fold lines if they get [????]

Freytag: No – take Ruby object as a clump and line break as a clump. Not trying to solve inline but trying to solve Ruby. If you import rich text into plain text, you can take out. At next level, you replace, xy & the anchor.

Becker: I think that you have to be careful on that. If you use, clarify and state whether annotating is visible or not. I haven't found a specification what I can and cannot do with this.

Freytag: There are some aspects that I don't know.

Joe: Not acceptable to say "I don't know what will happen." Need to write down.

Freytag: For the hidden *yomi* problem, I would like input.

Action Item [ ] for All UTC Members: Send information on hidden *yomi* to Freytag.

Action Item [ ] for Freytag: Revise for presentation at April UTC meeting.

Action Item [ ] for Aliprand/Winkler: Include on agenda of UTC #76/L2 #173 joint meeting:
Inline and interlinear annotations (Freytag) (schedule as late as possible in joint meeting.)

Whistler: Recommendation to give it a suggested code point of FF89B.

Last resort font
 [#75-M20] <u>Motion:</u> To publish Unicode Technical Report #12 documenting the last resort font as accepted by the UTC previously. "Last Resort Font" will be the preliminary title for UTR #12, but the title may be modified before publication.
Moved by Freytag, seconded by McGowan
Unanimous Approval

Action Item [   ] for Jenkins: Format text using template for UTRs, on etc. see Freytag.

Whistler: Have we thought through the consequences of "broken"cases?

Freytag: I have done a lot in discussion with Mark Davis. Add class of characters which will cause termination. Nesting will be allowed. Existing implementations of Ruby in Japan have built-in limitation.

6.   STANDING ITEMS
6.A. International Unicode Conference

7.   INTERACTION WITH SC2

7.A. WG2, WG3, and SC2 Plenary meetings in Seattle, WA
7.B. UTC positions on other topics for next WG2

Action Item [ ] for Freytag/Suignard: Submit final version of Criteria for disunification as input to the ISO/IEC 10646 procedures document. [#75-M7]

Action Item [ ] for Freytag/Suignard: Submit to WG2 for consideration: revised Western music proposal; revised Syriac proposal; proposal for Enclosing triangle; proposal for KIP currency sign.

Action Item [ ] for Freytag: As Unicode Liaison, submit UTC response to SC2/WG2 N1634.

7.C. IRG
7.C.1. Resolutions of IRG #10 [L2/98-016]

8.   REVIEW OF MOTIONS TO L2, AND ACTION ITEMS

Decision made that the next UTC/L2 joint meeting will be as scheduled on April 22-24th in Pennsylvania.

9.   CLOSING OF JOINT MEETING

Joan: Where do these go?
Action Item [ ] Aliprand: Announce BiDi ad hoc to be held in Seattle on Wednesday, March 18, beginning
at 3 pm.

Action Item [ ] Winkler: Check whether Alain La Bonté received the revised minutes of the disucssion on international string ordering.

Action Item [ ] for Freytag: Arrange for addition of all formatted additions to pipeline section of Web site.

APPENDIX I
UTC #75/L2#172 JOINT MEETING
FEBRUARY 25-27, 1998
CUPERTINO, CA
(alpha by company)

NAME	COMPANY	E-MAIL
Peter Edberg	Apple	pedberg@apple.com
Peter Lofting	Apple	lofting@apple.com
Rick McGowan	Apple	rmcgowan@apple.com
John Jenkins	Apple	jenkins@apple.com
Don Carroll	Hewlett-Packard	dcarroll@sea.hp.com
Mike Ksar	Hewlett-Packard	mike_ksar@hp.com
Mark Davis	IBM	mark@unicode.org
Lisa Moore	IBM	lisam@us.ibm.com
V.S. Umamaheswaran	IBM	umavs@ca.ibm.com
Brendan Murray	Lotus	brendan_murray@lotus.com
George Kiraz	Lucent	gkiraz@research.bell-labs.com
Paul Nelson	Microsoft	a-pauln@microsoft.com
Michel Suignard	Microsoft	michelsu@microsoft.com
Murray Sargent III	Microsoft	murrays@microsoft.com
Gary Roberts	NCR	gary.roberts@SanDiego.CA.NCR.com
Chris Boyle	Novell	dcboyle@novell.com
Lloyd Honomichi	Novell	lloyd_honomichi@novell.com
Yu Gong	Oracle	ygong@u.s.oracle.com
Nelson Ng	Oracle	nng@u.s.oracle.com
John Fiscella	Production First Software	profirst@compuserve.com
Tex Texin	Progress	texin@progress.com
Joan Aliprand	RLG	br.jma@rlg.org
Ed Hart	SHARE	edwin.hart@jhuopl.edu
Ken Whistler	Sybase	kenw@sybase.com
Julie Allen	Unicode	adollen@ix.netcom.com
Asmus Freytag	Unicode	asmus@unicode.org
Julia Oesterle	Unicode	julia@unicode.org
Arnold Winkler	Unisys	arnold.winkler@unisys.com
Joe Becker	Xerox	becker.osbu_north@xerox.com